Matrix Computations **Fall 2020**

Lecture 5: Eigenvalue Problems

Lecturer: Prof. Yue Qiu & Prof. Ziping Zhao *Scribe: Lin Zhu & Yijia Chang*

1 Basic Concepts

1.1 Eigenvalues & Eigenvectors

Matrices can be used to represent linear transformations. Their effects can be: rotation, reflection, translation, scaling, permutation, etc., and combinations thereof.

$$\mathbf{A} = \begin{bmatrix}\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$\mathbf{A}\mathbf{v} = \begin{bmatrix} v_1\mathbf{a}_1, v_2\mathbf{a}_2, v_3\mathbf{a}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1\mathbf{v} \\ \mathbf{b}_2\mathbf{v} \\ \mathbf{b}_3\mathbf{v} \end{bmatrix}$$

Example 1. *Let*

$$\mathbf{A} = \begin{bmatrix} 3 \\ 5 \end{bmatrix},$$

- **BA**: *rotate* **A** *for of θ degrees counterclockwise, where*

$$\mathbf{B} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix};$$

- **BA**: **A** *is stretched c times as far, where*

$$\mathbf{B} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

There is a much simple instance,

$$\mathbf{A} = \begin{bmatrix} 2 & & \\ & 3 & \\ & & 5 \end{bmatrix} \text{ then } \mathbf{A}\mathbf{v} = \begin{bmatrix} 2v_1 \\ 3v_2 \\ 5v_3 \end{bmatrix}.$$

Is there a set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ such that

$$\mathbf{v} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] \cdot \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3$$

and for any A,

$$\mathbf{A}\mathbf{v} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] \cdot \begin{bmatrix} \lambda_1 c_1 \\ \lambda_2 c_2 \\ \lambda_3 c_3 \end{bmatrix} = \lambda_1 \cdot c_1 \mathbf{x}_1 + \lambda_2 \cdot c_2 \mathbf{x}_2 + \lambda_3 \cdot c_3 \mathbf{x}_3.$$

These transformations can be rather complicated, and therefore we often want to decompose a transformation into a few simple actions that we can better understand. Finding singular values and associated singular vectors is one such approach.

A more basic approach is to consider eigenvalues and eigenvectors.

Definition 1. *For an $n \times n$ matrix $\mathbf{A}$, scalars λ and vectors $\mathbf{v}_{n\times 1} \neq \mathbf{0}$ satisfying $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ are called* **eigenvalues** *and* **eigenvectors** *of $\mathbf{A}$ associated with λ, respectively, and any such pair $(\lambda, \mathbf{v})$ is called an* **eigenpair** *for $\mathbf{A}$. The set of distinct eigenvalues of $\mathbf{A}$, denoted by $\sigma(\mathbf{A})$, is called the* **spectrum** *of $\mathbf{A}$.*

- $\lambda \in \sigma(\mathbf{A}) \iff \mathbf{A} - \lambda\mathbf{I}$ *is singular* $\iff \det(\mathbf{A} - \lambda\mathbf{I})$.
- $\{\mathbf{v} \neq \mathbf{0} \mid \mathbf{v} \in \mathcal{N}(\mathbf{A} - \lambda\mathbf{I})\}$ *is the set of all eigenvectors associates with λ. From now on, $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I})$ is called an* **eigenspaces** *for $\mathbf{A}$.*
- *Nonzero row vector $\mathbf{y}$ such that $\mathbf{y}(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$ is called a* **left-hand eigenvector** *of $\mathbf{A}$ associated with λ.*

Remark 1. *Eigenvectors specify the directions in which the matrix action is simple: any vector parallel to an eigenvector is changed only in length and/or orientation by the matrix $\mathbf{A}$.*

Definition 2. *The* **characteristic polynomial** *of $\mathbf{A}_{n\times n}$ is $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$. The* **characteristic equation** *for $\mathbf{A}$ is $p(\lambda) = 0$.*

Example 2. *Even real matrices can have complex eigenvalues. For instance,*

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

has a characteristic polynomial

$$p(\lambda) = \begin{vmatrix} \lambda & -1 \\ 1 & \lambda \end{vmatrix} = \lambda^2 + 1,$$

so its eigenvalues are $\lambda_1 = i$ and $\lambda_2 = -i$ with associated eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix} \text{ and } \mathbf{v}_2 = \begin{bmatrix} i \\ 1 \end{bmatrix},$$

and associated eigenspaces $\mathbf{V}_1 = \mathsf{span}(\mathbf{v}_1)$ and $\mathbf{V}_2 = \mathsf{span}(\mathbf{v}_2)$.

Some facts about eigenvalues:

1. If $\lambda^n + c_1\lambda^{n-1} + c_2\lambda^{n-2} + \ldots + c_{n-1}\lambda + c_n = 0$ is the characteristic equation for $\mathbf{A}_{n\times n}$, and its n eigenvalues are $\lambda_1, \lambda_2, \ldots, \lambda_n$, then

- $\text{trace}(\mathbf{A}) = \lambda_1 + \lambda_2 + \ldots + \lambda_n = -c_1$;
- $\det(\mathbf{A}) = \lambda_1 \lambda_2 \cdots \lambda_n = (-1)^n c_n$.

2. Eigenvalues of $\mathbf{A}^k$ are $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$.

3. Eigenvalues of real symmetric matrices are real.

4. Eigenvectors of real symmetric matrices are real.

5. Complex eigenvalues of real matrices appear in conjugate pair.

   - For $\mathbf{A} \in \mathbb{R}^{n \times n}$, if $(\lambda, \mathbf{v})$ is an eigenpair, then also $(\bar{\lambda}, \mathbf{v}^H)$.

6. Skew-Hermitian matrices ($\mathbf{A} = -\mathbf{A}^H$) have only pure imaginary eigenvalues.

Theorem 1. *The eigenvectors of distinct eigenvalues are linearly independent.*

Proof. For a $n \times n$ matrix $\mathbf{A}$, let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be its k distinct eigenvalues with $k \leq n$, let $\mathbf{v}_i$ be the corresponding eigenvectors associated with λ_i for all $i \in [k]$[1].

Assume that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly dependent and there is a maximal linearly independent set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_\ell\}$ with $\ell < k$. Then there exists a nonzero set $\{c_i \mid i \in [\ell]\}$ such that $\sum_{i=1}^{\ell} c_i \mathbf{v}_i = \mathbf{v}_{\ell+1}$. We then have that

$$\mathbf{A}\mathbf{v}_{\ell+1} = \sum_{i=1}^{\ell} c_i \cdot \mathbf{A}\mathbf{v}_i = \sum_{i=1}^{\ell} c_i \cdot \lambda_i \mathbf{v}_i,$$

$$\mathbf{A}\mathbf{v}_{\ell+1} = \lambda_{\ell+1}\mathbf{v}_{\ell+1} = \lambda_{\ell+1} \sum_{i=1}^{\ell} c_i \cdot \mathbf{v}_i.$$

Hence,

$$\sum_{i=1}^{\ell} (\lambda_i - \lambda_{\ell+1}) c_i \mathbf{v}_i = \mathbf{0},$$

which contradicts to $\lambda_i \neq \lambda_{\ell+1}$ for all $i \in [\ell]$. □

[1]For a positive integer n, denote by $[n]$ the set $\{1, 2, \ldots, n\}$.

1.2 Algebraic Multiplicity & Geometric Multiplicity

Considering such a matrix

$$\mathbf{M} = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

we can find that $\mathbf{M}$ has 2 distinct eigenvalues: 2 and 3, where 3 is a double root of the characteristic polynomial for $\mathbf{M}$. Now, if the eigenspace corresponding to 3 also has two basis vectors, this would not be strange, but instead the eigenspace corresponding to 3 is the span of only one vector $(0, 1, 0)^T$. This leads us to two definitions:

Definition 3. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix with eigenvalue* λ*. The* **algebraic multiplicity** *of* λ *is the number of times* λ *is repeated as a root of the characteristic polynomial* $p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$.

Definition 4. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix with eigenvalue* λ*. The* **geometric multiplicity** *of* λ *is the dimension of* $\mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$*, the eigenspace of* λ.

Example 3. *For the mentioned-above* $\mathbf{M}$*, the characteristic polynomial is* $p(\lambda) = \det(\mathbf{M} - \lambda \mathbf{I}) = (\lambda - 2)(\lambda - 3)^2$*, so* $\lambda_1 = 2$ *with algebraic multiplicity* $\mu_1 = 1$ *and* $\lambda_2 = 3$ *with algebraic multiplicity* $\mu_2 = 2$.

Since

$$\mathbf{M} - \lambda_1 \mathbf{I} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{M} - \lambda_2 \mathbf{I} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

we have $\dim(\mathcal{N}(\mathbf{M} - \lambda_1 \mathbf{I})) = 1$ *and* $\dim(\mathcal{N}(\mathbf{M} - \lambda_2 \mathbf{I})) = 1$*. Hence, the geometric multiplicity of* λ_1 *is* $\gamma_1 = 1 = \mu_1$ *and that of* λ_2 *is* $\gamma_2 = 1 < \mu_2$.

Theorem 2. *Let* λ_i *be an eigenvalue of the* $n \times n$ *matrix* $\mathbf{A}$*, then its algebraic multiplicity is at least as large as its geometric multiplicity.*

Proof. Refer to http://www.ee.iitm.ac.in/uday/2017b-EE5120/multiplicity.pdf or https://www.statlect.com/matrix-algebra/algebraic-and-geometric-multiplicity-of-eigenvalues. □

Definition 5. *Let* λ *be an eigenvalue of matrix* $\mathbf{A}$*, Then* λ *is called* **defective** *if its geometric multiplicity is less than its algebraic multiplicity.*

Definition 6. *A matrix* $\mathbf{A}$ *is called* **nondefective** *if it has no defective eigenvalue.*

1.3 Similarity Transformation & Diagonalization

Definition 7. *Consider two matrices* $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$*. Then* $\mathbf{A}$ *and* $\mathbf{B}$ *are similar if there exists a nonsingular matrix* $\mathbf{S} \in \mathbb{C}^{n\times n}$ *such that* $\mathbf{A} = \mathbf{S}^{-1}\mathbf{B}\mathbf{S}$*.*

Definition 8. *Two matrices* $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$ *are said to be unitarily similar if and only if there exists a unitary matrix* $P \in \mathbb{C}^{n\times n}$ *such that* $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$*.*

Remark 2. *If* $\mathbf{P}$ *is Hermitian, the similarity transformation can be written as* $\mathbf{A} = \mathbf{P}^H\mathbf{B}\mathbf{P}$*. If* $\mathbf{P}$ *is real, then the matrix is orthogonal,* $\mathbf{P}^{-1} = \mathbf{P}^T$ *and the similarity transformation becomes* $\mathbf{A} = \mathbf{P}^T\mathbf{B}\mathbf{P}$*.*

Theorem 3. (Similarity is an Equivalence Relation) *Suppose* $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{C}^{n\times n}$*. Then the following properties hold:*

- Reflexive: $\mathbf{A}$ *is similar to* $\mathbf{A}$*.*
- Symmetric: *If* $\mathbf{A}$ *is similar to* $\mathbf{B}$*, then* $\mathbf{B}$ *is similar to* $\mathbf{A}$*.*
- Transitive: *If* $\mathbf{A}$ *is similar to* $\mathbf{B}$ *and* $\mathbf{B}$ *is similar to* $\mathbf{C}$*, then* $\mathbf{A}$ *is similar to* $\mathbf{C}$*.*

Definition 9. *A matrix* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *is said to be* **diagonalizable** *if it is similar to a diagnoal matrix; i.e., there exists a nonsingular* $\mathbf{S} \in \mathbb{C}^{n\times n}$ *and diagonal* $\mathbf{\Lambda} \in \mathbb{C}^{n\times n}$ *such that*

$$\mathbf{\Lambda} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}, \tag{1}$$

or equivalently,

$$\mathbf{A} = \mathbf{S}\lambda_i\mathbf{S}^{-1}.$$

Remark 3. *The above (1) is equivalent to* $\mathbf{A}\mathbf{S} = \mathbf{S}\mathbf{\Lambda}$ *or*

$$\mathbf{A}\mathbf{s}_i = \lambda_i\mathbf{s}_i, \ \forall i \in [n],$$

where $\mathbf{s}_i$ *is the* i*-th column vector of* $\mathbf{S}$ *and* λ_i *is the* (i,i)*-th entry of* $\mathbf{\Lambda}$*. Hence, every* $(\mathbf{s}_i, \lambda_i)$ *is an eigenpair of* $\mathbf{A}$*.*

Theorem 4. *A matrix* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *is nondefective, i.e., has no defective eigenvalues, if and only if* $\mathbf{A}$ *is similiar to a diagonal matrix, i.e.,*

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1},$$

where $\mathbf{V} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ *is the matrix formed with the eigenvalues of* $\mathbf{A}$ *as its columns and* $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *contains the eigenvalues.*

Remark 4. *Nondefective matrices are* diagonalizable*. Nondefective matrices have linearly independent eigenvectors.*

Exercise 1:

1. The $\mathbf{A}$ is invertible $\iff$ 0 is $0 \notin \sigma(\mathbf{A})$.
2. If $\mathbf{A}$ is invertible, then λ^{-1} is an eigenvalue of $\mathbf{A}^{-1}$ if only if λ is an eigenvalue of $\mathbf{A}$.
3. If $p(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_k x^k$ is any polynomial, then we define $p(\mathbf{A})$ to be the matrix
$$p(\mathbf{A}) = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} + \alpha_2 \mathbf{A}^2 + \ldots + \alpha_k \mathbf{A}^k.$$
Show that if $(\lambda, \mathbf{v})$ is an eigenpair for $\mathbf{A}$, then $(p(\lambda), \mathbf{v})$ is an eigenpair for $p(\mathbf{A})$.
4. Similiar matrices always have the same characteristic polynomial.
5. Similiar matrices always have the same eigenvalues.

Solution:

1. The matrix $\mathbf{A}$ is invertible iff $\mathcal{N}(\mathbf{A} - 0 \cdot \mathbf{I}) = \{\mathbf{0}\}$ iff 0 is not an eigenvalue of $\mathbf{A}$.
2. If $\mathbf{A}$ is invertible, all eigenvalues of $\mathbf{A}$ and $\mathbf{A}^{-1}$ are nonzero by the last claim. For $\lambda \neq 0$, note that $\mathbf{A}\mathbf{v} = \lambda^{-1}\mathbf{v}$ iff $\mathbf{A}\mathbf{v} = \lambda \mathbf{A}(\lambda^{-1}\mathbf{v}) = \lambda \mathbf{v}$.
Thus, λ^{-1} is an eigenvalue of $\mathbf{A}^{-1}$ iff there is a nonzeor vector $\mathbf{v}$ such that $\mathbf{A}^{-1}\mathbf{v} = \lambda^{-1}\mathbf{v}$ iff there is a nonzero vector $\mathbf{v}$ such that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ iff λ is an eigenvalue of $\mathbf{A}$.
3. $$p(\mathbf{A})\mathbf{v} = (\sum_{i=0}^{k} \alpha_i \mathbf{A}^i)\mathbf{v} = \sum_{i=0}^{k} \alpha_i \mathbf{A}^i \mathbf{v} = \sum_{i=0}^{k} \alpha_i \lambda^i \mathbf{v} = (\sum_{i=0}^{k} \alpha_i \lambda^i)\mathbf{v} = p(\lambda)\mathbf{v}.$$
4. Let $\mathbf{A}$ and $\mathbf{B}$ be two similar matrices. Thus there is an invertible matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$. Thus the characteristic polynomial of $\mathbf{A}$ is
$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) &= \det(\mathbf{P}^{-1}\mathbf{B}\mathbf{P} - \lambda\mathbf{P}^{-1}\mathbf{I}\mathbf{P}) = \det(\mathbf{P}^{-1}(\mathbf{B} - \lambda\mathbf{P})\mathbf{P}) \\ &= \det(\mathbf{P}^{-1})\det(\mathbf{B} - \lambda\mathbf{I})\det(\mathbf{P}) = \det(\mathbf{B} - \lambda\mathbf{I})\det(\mathbf{P})\det(\mathbf{P}\mathbf{P}^{-1}) \\ &= \det(\mathbf{B} - \lambda\mathbf{I}) = \text{ the characteristic polynomial of } \mathbf{B}. \end{aligned}$$
5. If $\mathbf{A}$ and $\mathbf{B}$ are similar, there is some ivertible matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$. If $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, then we have $\mathbf{B}(\mathbf{P}\mathbf{v}) = \mathbf{P}(\mathbf{P}^{-1}\mathbf{B}\mathbf{P})\mathbf{v} = \mathbf{P}(\mathbf{A}\mathbf{v}) = \lambda\mathbf{P}\mathbf{v}$ and if $\mathbf{B}\mathbf{v} = \lambda\mathbf{v}$ we have $\mathbf{A}(\mathbf{P}^{-1}\mathbf{v}) = \lambda\mathbf{P}^{-1}\mathbf{v}$.

Exercise 2 (Ture or False):

1. Any $n \times n$ matrix has fewer than n eigenvalues is not diagonalizable.

2. Let λ be an eigenvalue of any $\mathbf{A}$ and $\mathbf{V}_\lambda$ be the associated eigenspace.
   (a) Every $\mathbf{v} \in \mathbf{V}_\lambda$ is an eigenvector of $\mathbf{A}$ associated with λ.
   (b) The matrix $\mathbf{A}$ is diagonalizable $\iff$ the multiplicity of λ equals the dimension of $\mathbf{V}_\lambda$.

3.
   - $\lambda \in \sigma(\mathbf{A})$ and $\mu \in \sigma(\mathbf{B}) \implies \lambda + \mu \in \sigma(\mathbf{A} + \mathbf{B})$.
   - $\lambda \in \sigma(\mathbf{A})$ and $\mu \in \sigma(\mathbf{B}) \implies \lambda\mu \in \sigma(\mathbf{AB})$.

Solution:

1. False. Take $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

2. (a) False. The zero vector is not.
   (b) False. It needs one more condition that all the eigenvalues of $\mathbf{A}$ is real. For instance, the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ has no real eigenvalue.

3. False. Take
$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

2 Decomposition

2.1 Eigendecomposition

Theorem 5. *Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be its eigenvalues. The matrix $\mathbf{A}$ is said to be diagonalizable, or admits an eigendecomposition, if there exists an invertible $\mathbf{V} \in \mathbb{C}^{n\times n}$ such that*

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1},$$

where $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$.

Remark 5. *Eigendecomposition is also known as eigenvalue decomposition or spectral decomposition.*

Theorem 6. (**the sufficient condition for the existence of eigendecomposition**) *For $\mathbf{A} \in \mathbb{C}^{n\times n}$, it admits an eigendecomposition if it has n distinct eigenvalues.*

Theorem 7. (**the sufficient and necessary condition for the existence of eigendecomposition**) *For $\mathbf{A} \in \mathbb{C}^{n\times n}$, it admits an eigendecomposition if and only if for all the eigenvalue, its algebraic multiplicity equals the geometric multiplicity.*

Theorem 8. (**the sufficient and necessary condition for the existence of eigendecomposition**) *For $\mathbf{A} \in \mathbb{C}^{n\times n}$, it admits an eigendecomposition if and only if it has n linear independent eigenvectors.*

2.2 Eigendecomposition for Hermitian & Real Symmetric Matrices

Let $\mathbf{A} \in \mathbb{H}^{n\times n}$, then

- the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$ are real
- suppose that λ_i's are orders such that $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ is the set of all distinct eigenvalues of $\mathbf{A}$. Also, let $\mathbf{v}_i$ be *any* eigenvector associated with λ_i. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ must be orthonormal.

Theorem 9. *Every $\mathbf{A} \in \mathbb{H}^n$ admits an eigendecomposition*

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H,$$

where $\mathbf{V} \in \mathbb{C}^{n\times n}$ is unitary ($\mathbf{V}\mathbf{V}^H = \mathbf{I}$), $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ with $\lambda_i \in \mathbb{R}$ for all i. Also, if $\mathbf{A} \in \mathbb{S}^n$, $\mathbf{V}$ is orthogonal.

Proof. Let $\mathbf{A}$ be Hermitian and let $\mathbf{A} = \mathbf{V}\mathbf{T}\mathbf{V}^H$ be its Schur decomposition. As

$$\mathbf{0} = \mathbf{A} - \mathbf{A}^H = \mathbf{V}\mathbf{T}\mathbf{V}^H - \mathbf{V}\mathbf{T}^H\mathbf{V}^H = \mathbf{V}(\mathbf{T} - \mathbf{T}^H)\mathbf{V}^H \iff \mathbf{0} = \mathbf{T} - \mathbf{T}^H.$$

Since $\mathbf{T}$ is upper triangular and $\mathbf{T}^H$ is lower triangular, $\mathbf{T} = \mathbf{T}^H$ implies that $\mathbf{T}$ is diagonal. Thus, the Schur decomposition is also the eigendecomposition. $\square$

Note: $\mathbf{T} = \mathbf{T}^H$ also implies that $\mathbf{T}_{ii}$'s are real, so the proof also confirms that λ_i's are real.

Remark 6. *The above results apply to real symmetric matrices since $\mathbf{A} = \mathbf{A}^T \implies \mathbf{A} = \mathbf{A}^H$.*

Corollary 9.1. *If $\mathbf{A}$ is Hermitian or real symmetric, $\mu_i = \gamma_i$ for all λ_i (no. of repeated eigenvalues = no. of linearly independent eigenvectors)*

2.3 Schur Decomposition

Every square matrix is unitarily similar to an upper-triangular matrix.

Theorem 10. *Let* $\mathbf{A} \in \mathbb{C}^{n\times n}$ *and let* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *be its eigenvalues. The matrix* $\mathbf{A}$ *admits a decomposition*

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H,$$

for some unitary $\mathbf{U} \in \mathbb{C}^{n\times n}$ *and for some upper triangular* $\mathbf{T} \in \mathbb{C}^{n\times n}$ *with* $\mathbf{T}_{ii} = \lambda_i$.

Proof. Use induction on n, the size of the matrix. For $n = 1$, there is nothing to prove. For $n > 1$, assume that all $n-1 \times n-1$ matrices are unitarily similar to an upper-triangular matrix, and consider an $n \times n$ matrix $\mathbf{A}$. Suppose that $(\lambda, \mathbf{x})$ is an eigenpair for $\mathbf{A}$, and suppose that $\mathbf{x}$ has been normalized so that $\|\mathbf{x}\|_2 = 1$. As discussed on p. 325, we can construct an elementary reflector $\mathbf{R} = \mathbf{R}^* = \mathbf{R}^{-1}$ with the property that $\mathbf{R}\mathbf{x} = \mathbf{e}_1$ or, equivalently, $\mathbf{x} = \mathbf{R}\mathbf{e}_1$ (set $\mathbf{R} = \mathbf{I}$ if $\mathbf{x} = \mathbf{e}_1$). Thus $\mathbf{x}$ is the first column in $\mathbf{R}$, so $\mathbf{R} = \big(\mathbf{x} \,|\, \mathbf{V}\big)$, and

$$\mathbf{R}\mathbf{A}\mathbf{R} = \mathbf{R}\mathbf{A}\big(\mathbf{x} \,|\, \mathbf{V}\big) = \mathbf{R}\big(\lambda\mathbf{x} \,|\, \mathbf{A}\mathbf{V}\big) = \big(\lambda\mathbf{e}_1 \,|\, \mathbf{R}\mathbf{A}\mathbf{V}\big) = \begin{pmatrix} \lambda & \mathbf{x}^*\mathbf{A}\mathbf{V} \\ \mathbf{0} & \mathbf{V}^*\mathbf{A}\mathbf{V} \end{pmatrix}.$$

Since $\mathbf{V}^*\mathbf{A}\mathbf{V}$ is $n-1 \times n-1$, the induction hypothesis insures that there exists a unitary matrix $\mathbf{Q}$ such that $\mathbf{Q}^*(\mathbf{V}^*\mathbf{A}\mathbf{V})\mathbf{Q} = \tilde{\mathbf{T}}$ is upper triangular. If $\mathbf{U} = \mathbf{R}\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}$, then $\mathbf{U}$ is unitary (because $\mathbf{U}^* = \mathbf{U}^{-1}$), and

$$\mathbf{U}^*\mathbf{A}\mathbf{U} = \begin{pmatrix} \lambda & \mathbf{x}^*\mathbf{A}\mathbf{V}\mathbf{Q} \\ \mathbf{0} & \mathbf{Q}^*\mathbf{V}^*\mathbf{A}\mathbf{V}\mathbf{Q} \end{pmatrix} = \begin{pmatrix} \lambda & \mathbf{x}^*\mathbf{A}\mathbf{V}\mathbf{Q} \\ \mathbf{0} & \tilde{\mathbf{T}} \end{pmatrix} = \mathbf{T}$$

is upper triangular. Since similar matrices have the same eigenvalues, and since the eigenvalues of a triangular matrix are its diagonal entries (Exercise 7.1.3), the diagonal entries of $\mathbf{T}$ must be the eigenvalues of $\mathbf{A}$. ∎

2.4 Some facts:

1. $\mathbf{A} \in \mathbb{C}^{n\times n}$ is diagonalizable, i.e., $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, if and only if $\mathbf{A}$ is nondefective (i.e., $\mu_i = \lambda_i$).

2. $\mathbf{A} \in \mathbb{C}^{n\times n}$ is unitarily diagonalizable, i.e., $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, if and only if $\mathbf{A}$ is normal (i.e., $\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}$) (including Hermitian ($\mathbf{A}^H = \mathbf{A}$) and skew-Hermitian matrices ($\mathbf{A}^H = -\mathbf{A}$)).

3. $\mathbf{A} \in \mathbb{C}^{n\times n}$ is unitarily triangularizable, i.e., $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ with unitary $\mathbf{U}$ for any $\mathbf{A}$.

4. $\mathbf{A} \in \mathbb{C}^{n\times n}$ has Jordan canonical/normal from, i.e., $\mathbf{A} = \mathbf{S}\mathbf{J}\mathbf{S}^{-1}$ with Jordan block $\mathbf{J}$ for any $\mathbf{A}$, where

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ & & & \mathbf{J}_k \end{bmatrix} \text{ with a square } \mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}.$$

Exercise 3:

1. $\mathsf{rank}(\mathbf{A}) \geq$ number of nonzero eigenvalues of $\mathbf{A}$.
2. If $\mathbf{A}$ is Hermitian, then all of eigenvalues of $\mathbf{A}$ are real.
3. If $\mathbf{A}$ admits an eigendecomposition (eigenvalue decomposition), $\mathsf{rank}(\mathbf{A})$ = number of nonzero eigenvalues of $\mathbf{A}$.
4. For any $\mathbf{A} \in \mathbb{H}^n$, and any $\mathbf{x} \in \mathbb{C}^n$,

$$\lambda_1 \leq \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \lambda_n,$$

where λ_1 and λ_n are the smallest and largest eigenvalue, respectively.

Solution:

1. see HW5-P2
2. see HW5-P2
3. see HW5-P2
4. Denote by $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ the eigendecomposition of $\mathbf{A}$, where $\mathbf{v} \in \mathbb{C}^{n\times n}$ is unitary. Then

$$\|\mathbf{V}\mathbf{x}\|_2 = (\mathbf{V}\mathbf{x})^T(\mathbf{V}\mathbf{x}) = \mathbf{x}^T\mathbf{V}^T\mathbf{V}\mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|_2 \implies \|\mathbf{V}\mathbf{x}\|_2 = \|\mathbf{x}\|_2.$$

Consequently,

$$\frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|\mathbf{V}\mathbf{\Lambda}\mathbf{V}^H\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|\mathbf{\Lambda}\mathbf{V}^H\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|\mathbf{\Lambda}\mathbf{V}^H\mathbf{x}\|_2}{\|\mathbf{V}^H\mathbf{x}\|_2} = \frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$$

with $y = \mathbf{V}^H\mathbf{x}$.

Hence,

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \min_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \quad \text{and} \quad \min_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \leq \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \max_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2}.$$

As,

$$\frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} = \sqrt{\frac{\sum_{i=1}^n \lambda_i^2\mathbf{x}_i^2}{\sum_{i=1}^n \mathbf{x}_i^2}} \leq \sqrt{\frac{\sum_{i=1}^n \lambda_n^2\mathbf{x}_i^2}{\sum_{i=1}^n \mathbf{x}_i^2}} = \lambda_n,$$
$$\frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} = \sqrt{\frac{\sum_{i=1}^n \lambda_i^2\mathbf{x}_i^2}{\sum_{i=1}^n \mathbf{x}_i^2}} \geq \sqrt{\frac{\sum_{i=1}^n \lambda_1^2\mathbf{x}_i^2}{\sum_{i=1}^n \mathbf{x}_i^2}} = \lambda_1,$$

and $\frac{\|\mathbf{\Sigma}\mathbf{e}_1\|_2}{\|\mathbf{e}_1\|_2} = \lambda_1$, $\frac{\|\mathbf{\Sigma}\mathbf{e}_n\|_2}{\|\mathbf{e}_n\|_2} = \lambda_n$, we have

$$\lambda_1 \leq \min_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \text{ and } \max_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \leq \lambda_n.$$

Therefore,

$$\lambda_1 \leq \min_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \leq \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \max_{\mathbf{y}\neq\mathbf{0}} \frac{\|\mathbf{\Lambda}\mathbf{y}\|_2}{\|\mathbf{y}\|_2} \leq \lambda_n.$$

3 Variational Characterization of Eigenvalues of Hermitian Matrices

In this part, to obtain the eigenvalues of Hermitian matrices, variational characterization of eigenvalues is introduced based on the concept of "Rayleigh quotient".

Definition 10. *For any* $\mathbf{x} \in \mathbb{C}^n$ *with* $\mathbf{x} \neq \mathbf{0}$, *the ratio*

$$R(\mathbf{x}) = \frac{\mathbf{x}^H \mathbf{A} \mathbf{x}}{\mathbf{x}^H \mathbf{x}}$$

is called the Rayleigh quotient.

Theorem 11. ***(Courant-Fischer Minimax Theorem)*** *Let* $\mathbf{A} \in \mathbb{C}^{n \times n}$ *be a Hermitian matrix,* $\mathcal{S}_k$ *denote a subspace of* $\mathbb{C}^n$ *of dimension* k, *and* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ *represent the eigenvalues of* $\mathbf{A}$. *For any* $k \in \{1, 2, \ldots, n\}$, *it holds that*

$$\lambda_k = \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{C}^n} \max_{\mathbf{x} \in \mathcal{S}_{n-k+1}, \|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x} = \max_{\mathcal{S}_k \subseteq \mathbb{C}^n} \min_{\mathbf{x} \in \mathcal{S}_k, \|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x}.$$

Proof: Let $u_1, u_2, \ldots, u_n$ be orthonormal eigenvectors of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively.

(Proof of min-max form) Let $\mathcal{V}$ denote a subspace of $\mathbb{R}^n$ of dimension $n - k + 1$. Let $\mathcal{W}$ denote the subspace spanned by $\{u_1, u_2, \ldots, u_k\}$. We can observe that $\dim(\mathcal{V} \cap \mathcal{W}) \geq 1$, which implies that $\mathcal{V} \cap \mathcal{W} \neq \{\mathbf{0}\}$.

Consider such a vector $\mathbf{a} \in \mathcal{V} \cap \mathcal{W}$ that $\|\mathbf{a}\|_2 = 1$. Since $\mathbf{a} \in \mathcal{W}$, we can derive that $\mathbf{a} = \sum_{i=1}^{k} c_i u_i$ with $\|\mathbf{a}\|_2 = 1$, and then

$$\mathbf{a}^T \mathbf{A} \mathbf{a} = \sum_{i=1}^{k} \lambda_i c_i^2 \geq \lambda_k \sum_{i=1}^{k} c_i^2 = \lambda_k,$$

which implies that $\lambda_k \leq \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{R}^n} \max_{\mathbf{x} \in \mathcal{S}_{n-k+1}, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A} \mathbf{x}$.

Taking a special $\mathcal{V}$ as the subspace spanned by $\{u_k, u_{k+1}, \ldots, u_n\}$, we can get that $u_k \in \mathcal{V} \cap \mathcal{W}$ and

$$u_k^T \mathbf{A} u_k = \lambda_k u_k^T \mathbf{A} u_k = \lambda_k.$$

Hence, the claim is proved that

$$\lambda_k = \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{R}^n} \max_{\mathbf{x} \in \mathcal{S}_{n-k+1}, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A} \mathbf{x}.$$

(Proof of max-min form) Let $\mathcal{V}$ denote a subspace of $\mathbb{R}^n$ of dimension k. Let $\mathcal{W}$ denote the subspace spanned by $\{u_k, u_{k+1}, \ldots, u_n\}$. We can observe that $\dim(\mathcal{V} \cap \mathcal{W}) \geq 1$, which implies that $\mathcal{V} \cap \mathcal{W} \neq \{\mathbf{0}\}$.

Consider such a vector $\mathbf{a} \in \mathcal{V} \cap \mathcal{W}$ that $\|\mathbf{a}\|_2 = 1$. Since $\mathbf{a} \in \mathcal{W}$, we can derive that $\mathbf{a} = \sum_{i=k}^{n} c_i u_i$ with $\|\mathbf{a}\|_2 = 1$, and then

$$\mathbf{a}^T \mathbf{A} \mathbf{a} = \sum_{i=k}^{n} \lambda_i c_i^2 \leq \lambda_k \sum_{i=k}^{n} c_i^2 = \lambda_k,$$

which implies that $\lambda_k \geq \max_{\mathcal{S}_k \subseteq \mathbb{C}^n} \min_{\mathbf{x} \in \mathcal{S}_k, \|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x}$.

Taking a special $\mathcal{V}$ as the subspace spanned by $\{u_1, u_2, \ldots, u_k\}$, we can get that $u_k \in \mathcal{V} \cap \mathcal{W}$ and

$$u_k^T \mathbf{A} u_k = \lambda_k u_k^T u_k = \lambda_k.$$

Hence, the claim is proved that

$$\lambda_k = \max_{\mathcal{S}_k \subseteq \mathbb{C}^n} \min_{\mathbf{x} \in \mathcal{S}_k, \|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x}.$$

Next, we introduce several theorems and corollaries that can be considered as consequences of the Courant-Fischer's theorem. The first Rayleigh-Ritz theorem is about the largest and the smallest eigenvalues of a Hermitian matrix, which can be directly derieved from the Courant-Fischer's theorem.

Theorem 12. ***(Rayleigh-Ritz Theorem)*** *Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a Hermitian matrix and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ represent the eigenvalues of $\mathbf{A}$. Then we have*

$$\lambda_1 = \max_{\|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x}$$

and

$$\lambda_n = \min_{\|\mathbf{x}\|_2 = 1} \mathbf{x}^H \mathbf{A} \mathbf{x}$$

The second theorem, by Weyl, allows us to obtain a lower and upper bound for the kth eigenvalue of $\mathbf{A} + \mathbf{B}$.

Theorem 13. ***(Weyl Theorem)*** *For any Hermitian matrix $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n \times n}$ and vector $\boldsymbol{z} \in \mathbb{C}^n$,*

$$\lambda_k(\mathbf{A}) + \lambda_n(\mathbf{B}) \leq \lambda_k(\mathbf{A} + \mathbf{B}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B})$$

for $k = 1, 2, 3, \ldots, n$.

Proof: By Courant-Fischer's theorem, for any $1 \leq k \leq n$,

$$\lambda_k(\mathbf{A} + \mathbf{B}) = \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{R}^n} \max_{\mathbf{x} \in \mathcal{S}_{n-k+1}, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T (\mathbf{A} + \mathbf{B}) \mathbf{x} = \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{R}^n} \max_{\mathbf{x} \in \mathcal{S}_{n-k+1}, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{B} \mathbf{x}$$

Note that by Rayleigh-Ritz's theorem, we know that

$$\mathbf{x}^T \mathbf{B} \mathbf{x} \geq \lambda_n(\mathbf{B}).$$

Hence, we can prove that

$$\lambda_k(\mathbf{A} + \mathbf{B}) \geq \lambda_n(\mathbf{B}) + \min_{\mathcal{S}_{n-k+1} \subseteq \mathbb{R}^n} \max_{\mathbf{x} \in \mathcal{S}_{n-k+1}, \|\mathbf{x}\|_2 = 1} \mathbf{x}^T \mathbf{A} \mathbf{x} = \lambda_n(\mathbf{B}) + \lambda_k(\mathbf{A})$$

The inequality

$$\lambda_k(\mathbf{A} + \mathbf{B}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B})$$

can be proved in the similar way.

The following interlacing theorem show the eigenvalues of a matrix derived from a matrix $\mathbf{A}$ relatively to the eigenvalues of $\mathbf{A}$.

Theorem 14. ***(Interlacing Theorem)*** *Let* $\mathbf{A}$ *be an* $n \times n$ *Hermitian matrix and* $\mathbf{B}$ *be an* $m \times m$ *principal sub-matrix of* $\mathbf{A}$ *(obtained by deleting both i-th row and i-th column for some values of i). Suppose* $\mathbf{A}$ *has eigenvalues* $\lambda_1 \le \ldots \le \lambda_n$ *and* $\mathbf{B}$ *has eigenvalues* $\beta_1 \le \ldots \le \beta_m$*. Then*

$$\lambda_k \le \beta_k \le \lambda_{k+n-m}, \qquad k = 1, \ldots, m$$

And if $m = n-1$*,*

$$\lambda_1 \le \beta_1 \le \lambda_2 \le \beta_2 \le \ldots \le \beta_{n-1} \le \lambda_n$$

Proof: Without loss of generality, assume $\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{X}^T \\ \mathbf{X} & \boldsymbol{Z} \end{bmatrix}$. Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be eigenvectors of $\mathbf{A}$, $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m\}$ be eigenvectors of $\mathbf{B}$. We define the following vector spaces:

$$V = span(\mathbf{x}_k, \ldots, \mathbf{x}_n); \quad W = span(\mathbf{y}_1, \ldots, \mathbf{y}_k); \quad \widetilde{W} = \{\begin{pmatrix} w \\ 0 \end{pmatrix} \in \mathbb{R}^n, w \in W\}$$

Since $dim(V) = n-k+1$ and $dim(\widetilde{W}) = dim(W) = k$, there exists $\widetilde{w} \in V \cap \widetilde{W}$ and $\widetilde{w} = \begin{pmatrix} w \\ 0 \end{pmatrix}$ for some $w \in W$ with $\|w\|_2 = 1$. Then

$$\widetilde{w}^T A \widetilde{w} = \begin{bmatrix} w^T & 0 \end{bmatrix} \begin{bmatrix} B & X^T \\ X & Z \end{bmatrix} \begin{bmatrix} w \\ 0 \end{bmatrix} = w^T B w.$$

Recall $\lambda_k = \min_{x \in V} \frac{x^T A x}{x^T x}$ and $\beta_k = \max_{x \in W} \frac{x^T B x}{x^T x}$. Then we see that

$$\lambda_k \le \frac{\widetilde{w}^T A \widetilde{w}}{\widetilde{w}^T \widetilde{w}} = \frac{w^T B w}{w^T w} \le \beta_k.$$

The proof of the other inequality is similar. We now define the vector spaces as

$$V = span(\mathbf{x}_1, \ldots, \mathbf{x}_{k+n-m}); \quad W = span(\mathbf{y}_k, \ldots, \mathbf{y}_m); \quad \widetilde{W} = \{\begin{pmatrix} w \\ 0 \end{pmatrix} \in \mathbb{R}^n, w \in W\}$$

Since $dim(V) = k+n-m$ and $dim(\widetilde{W}) = dim(W) = m-k+1$, there exists $\widetilde{w} \in V \cap \widetilde{W}$ and $\widetilde{w} = \begin{pmatrix} w \\ 0 \end{pmatrix}$ for some $w \in W$ with $\|w\|_2 = 1$. Then

$$\widetilde{w}^T A \widetilde{w} = \begin{bmatrix} w^T & 0 \end{bmatrix} \begin{bmatrix} B & X^T \\ X & Z \end{bmatrix} \begin{bmatrix} w \\ 0 \end{bmatrix} = w^T B w.$$

Recall $\lambda_{k+n-m} = \max_{x \in V} \frac{x^T A x}{x^T x}$ and $\beta_k = \min_{x \in W} \frac{x^T B x}{x^T x}$. Then we see that

$$\lambda_{k+n-m} \ge \frac{\widetilde{w}^T A \widetilde{w}}{\widetilde{w}^T \widetilde{w}} = \frac{w^T B w}{w^T w} \ge \beta_k.$$

References

[1] Taboga, Marcos. "Similar matrix", *https://www.statlect.com/matrix-algebra/similar-matrix*.

[2] Meyer, Carl D. Matrix analysis and applied linear algebra. Siam, 2000.