IFT 6760A - Lecture 4
Linear Algebra Refresher

Scribe(s): Michel Ma, Shenyang Huang
(adapted from Arthur Dehgan, Adrien Mainka last year's notes) **Instructor:** Guillaume Rabusseau

1 Summary

In the previous lecture, we discussed linear regression and its connection to orthogonal projection. We also introduced various forms of Singular Value Decomposition or SVD. Then, we talked about matrix inverse and determinant as well as how those are related to Moore-Penrose pseudo-inverse. Lastly, we gave definition for eigenvalues and eigenvectors and saw how they relate to matrix diagonalization.

In this lecture we will go more in depth on eigenvalues and eigenvectors and then introduce the Spectral Theorem and its connection to diagonalizable matrices. We will also see the Schur Decomposition and Jordan Canonical Form.

2 Eigenvalues and eigenvectors (continued)

Definition 1 (Eigenvalues multiplicities)**.** *Let λ be an eigenvalue of $\mathbf{A}$ then :*

- *The* geometric multiplicity*, $m_g(\lambda)$ is the dimension of E_λ where $E_\lambda = \{\mathbf{v} \mid \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}$*
- *The* algebric multiplicity *$m_a(\lambda)$, is the multiplicity of λ as root of $\det(\mathbf{A} - \lambda\mathbf{I})$ (the characteristic polynomial of $\mathbf{A}$). Equivalently, $m_a(\lambda)$ is the dimension of the generalized eigenspace $\bigcup_{k\geq 0} \mathcal{N}((\mathbf{A} - \lambda\mathbf{I})^k)$*
  where $\bigcup_{k\geq 0} \mathcal{N}((\mathbf{A} - \lambda\mathbf{I})^k) = \mathcal{N}(\mathbf{I}) \cup \mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) \cup \mathcal{N}((\mathbf{A} - \lambda\mathbf{I})^2) \cup \ldots \cup \mathcal{N}((\mathbf{A} - \lambda\mathbf{I})^k)$

Observe that :

- $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) = E_\lambda$.

  Proof. By definition, $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) = \{\mathbf{v} \mid (\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0\} = \{\mathbf{v} \mid \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\} = E_\lambda$ □

- $m_g(\lambda) \leq m_a(\lambda)$.

  Proof. As seen above, $m_a(\lambda)$ is the dimension of the generalized eigenspace $\bigcup_{k\geq 0} \mathcal{N}((\mathbf{A} - \lambda\mathbf{I})^k)$ which includes $\mathcal{N}(\mathbf{A} - \lambda\mathbf{I}) = E_\lambda$ (the dimension of $m_g(\lambda)$). Thus, we have $m_g(\lambda) \leq m_a(\lambda)$. □

- If the geometric multiplicity is strictly smaller than the algebraic multiplicity of an eigenvalue of $\mathbf{A} \in \mathbb{R}^{n\times n}$, $m_g(\lambda) < m_a(\lambda)$, then $\mathbf{A}$ is not diagonalizable.

  Proof. Let $\lambda_1, \cdots, \lambda_k$ be the eigenvalues of $\mathbf{A}$. Suppose $m_g(\lambda_1) < m_a(\lambda_1)$. Then, it follows that the eigenspace E_1 has dimension strictly less than $m_a(\lambda_1)$, and consequently the dimension of $\cup_{i=1}^k E_k$ is strictly less than n and there cannot exists a basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$. □

Example 2. *Let* $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$*, the corresponding characteristic polynomial is* $(2-\lambda)^2$*. The eigenvalue of* $\mathbf{A}$ *is 2 with multiplicity* $m_a(\lambda) = 2$*.*

$$\begin{aligned} E_\lambda &= \mathcal{N}(\mathbf{A} - 2\mathbf{I}) \\ &= \mathcal{N}(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}) \\ &= span(\begin{bmatrix} 1 \\ 0 \end{bmatrix}) \end{aligned} \tag{1}$$

We have $m_g(\lambda) = dim(E_\lambda) = 1$ *which is different from its eigenvalue multiplicity* $m_a(\lambda) = 2$ *hence the matrix* $\mathbf{A}$ *is not diagonalizable.*

Property 3. *The eigenvectors* $\mathbf{v_1}, ..., \mathbf{v_n}$ *associated with* distinct *eigenvalues of* $\mathbf{A}$ *are linearly independent.*

Proof. Let $\mathbf{v_1}$ and $\mathbf{v_2}$ two eigenvectors of $\mathbf{A}$ associated with the distinct eigenvalues λ_1 and λ_2 and let a and b such that $a\mathbf{v_1} + b\mathbf{v_2} = 0$. It follows that :

$$\begin{aligned} \mathbf{A}(a\mathbf{v_1} + b\mathbf{v_2}) &= 0 \\ a\mathbf{A}\mathbf{v_1} + b\mathbf{A}\mathbf{v_2} &= 0 \\ a\lambda_1\mathbf{v_1} + b\lambda_2\mathbf{v_2} &= 0 \end{aligned} \tag{2}$$

But by multiplying $a\mathbf{v_1} + b\mathbf{v_2} = 0$ by λ_2, we also have :

$$a\lambda_2\mathbf{v_1} + b\lambda_2\mathbf{v_2} = 0 \tag{3}$$

Hence by substracting (3) and (4) we have :

$$\begin{aligned} (a\lambda_1\mathbf{v_1} + b\lambda_2\mathbf{v_2}) - (a\lambda_2\mathbf{v_1} + b\lambda_2\mathbf{v_2}) &= 0 \\ a(\lambda_1 - \lambda_2) &= 0 \end{aligned} \tag{4}$$

If the eigenvalues λ_1 and λ_2 are distinct, i.e, $\lambda_1 \neq \lambda_2$, we can conclude $a = 0$. Similar conclusion can be drawn for $b = 0$. Therefore the eigenvectors $\mathbf{v_1}$ and $\mathbf{v_2}$ are linearly independent.
We can then repeat the process for each pair of eigenvectors of $\mathbf{A}$ and we would conclude that the eigenvectors $\mathbf{v_1}, ..., \mathbf{v_n}$ associated with *distinct* eigenvalues of $\mathbf{A}$ are linearly independent. □

3 Spectral Theorem

Definition 4 (Positive definite and semi-definite matrices)**.** *A symmetric matrix* $\mathbf{A} \in \mathbb{R}^{m \times m}$ *is :*

- positive definite *if and only if all of its eigenvalues are strictly positive.*
- positive semi-definite *or* non-negative definite *if all of its eigenvalues are non-negative.*

Property 5. *For any matrix* $\mathbf{A} \in \mathbb{R}^{m \times n}$*,* $\mathbf{A}^\top\mathbf{A}$ *is positive semi-definite*

Proof. Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$. Let $\mathbf{v}$ be an eigenvector of $\mathbf{A^T A}$, and λ the corresponding eigenvalue. Therefore, we have

$$\begin{aligned} \mathbf{A}^T\mathbf{A}\mathbf{v} &= \lambda\mathbf{v} \\ \mathbf{v}^T\mathbf{A}^T\mathbf{A}\mathbf{v} &= \mathbf{v}^T\lambda\mathbf{v} \\ \mathbf{v}^T\mathbf{A}^T\mathbf{A}\mathbf{v} &= \lambda||\mathbf{v}||^2 \end{aligned} \tag{5}$$

Since $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

$$\begin{aligned}(\mathbf{Av})^T\mathbf{Av} &= \lambda\|\mathbf{v}\|^2 \\ \|\mathbf{Av}\|^2 &= \lambda\|\mathbf{v}\|^2\end{aligned} \tag{6}$$

We also have

$$\|\mathbf{Av}\|^2 \geq 0 \text{ and } \|\mathbf{v}\|^2 \geq 0 \tag{7}$$

Therefore, $\lambda \geq 0$ □

Theorem 6 (Spectral theorem). *If $\mathbf{A} \in \mathbb{R}^{m\times m}$ is symmetric then:*

(i) All of its eigenvalues are real

(ii) $\mathbf{A}$ is diagonalizable by an orthogonal matrix meaning that there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{m\times m}$ such that $\mathbf{A} = \mathbf{UDU}^\top$ with $\mathbf{D} \in \mathbb{R}^{m\times m}$ being diagonal

Proof.

(i) Let $\lambda \in \mathbb{C}$ be an eigenvalue and $\mathbf{v} \neq 0$, a corresponding eigenvector of $\mathbf{A}$. we denote $\overline{\mathbf{v}}$ by the complex conjugate of $\mathbf{v}$.

$$\mathbf{v}^\top\mathbf{A}\overline{\mathbf{v}} = \mathbf{v}^\top\mathbf{A}^\top\overline{\mathbf{v}} = (\mathbf{Av})^\top\overline{\mathbf{v}} = \lambda\mathbf{v}^\top\overline{\mathbf{v}} \tag{8}$$

Since $\mathbf{A}$ is real and symmetric, hence we have $\mathbf{A} = \overline{\mathbf{A}}$ and $\mathbf{A}^\top = \mathbf{A}$. We also have :

$$\mathbf{v}^\top\mathbf{A}\overline{\mathbf{v}} = \mathbf{v}^\top \cdot \overline{\mathbf{A}} \cdot \overline{\mathbf{v}} = \mathbf{v}^\top\overline{\mathbf{Av}} = \mathbf{v}^\top\overline{\lambda\mathbf{v}} = \overline{\lambda}\mathbf{v}^\top\overline{\mathbf{v}} \tag{9}$$

Now we have $\lambda\mathbf{v}^\top\overline{\mathbf{v}} = \overline{\lambda}\mathbf{v}^\top\overline{\mathbf{v}}$ and since $\mathbf{v}^\top\overline{\mathbf{v}} \neq 0$, then $\lambda = \overline{\lambda}$ hence $\lambda \in \mathbb{R}$.

(ii) Let $\mathbf{v} \neq 0$ a unit eigenvector for λ, i.e. $\mathbf{Av} = \lambda\mathbf{v}$, and $\mathbf{v}^\top\mathbf{v} = 1$. Let $\mathcal{U} = span(\mathbf{v})^\perp = \{\mathbf{u} \in \mathbb{R}^m | \langle \mathbf{u}, \mathbf{v}\rangle = 0\}$. We observe that $\mathcal{U}$ is $\mathbf{A}$ invariant, i.e. if $\mathbf{u} \in \mathcal{U}$, then $\mathbf{Au} \in \mathcal{U}$. $\langle \mathbf{Au}, \mathbf{v}\rangle = (\mathbf{Au})^\top\mathbf{v} = \mathbf{u}^\top\mathbf{A}^\top\mathbf{v} = \mathbf{u}^\top\mathbf{Av} = \lambda\mathbf{u}^\top\mathbf{v} = 0$ which implies that we have $\mathbf{Au} \in \mathcal{U}$.

Let $\mathbf{u_1}, ..., \mathbf{u_{m-1}} \in \mathbb{R}^m$ be an orthonormal basis of $\mathcal{U}$, and let $\mathbf{U} = \begin{pmatrix} | & | & & | \\ \mathbf{v} & \mathbf{u_1} & \cdots & \mathbf{u_{m-1}} \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{m\times m}$.

We claim that $\mathbf{U}$ is orthogonal and $\mathbb{R}^m = \mathcal{U}\oplus\mathcal{U}^\top = \mathcal{U}\oplus span(\mathbf{v})$, with $dim(span(\mathbf{v})) = 1$. We first prove that $\mathbf{U}$ is orthogonal since $\mathcal{U} = span(\mathbf{v})^\perp$. Let's look at the first column of $\mathbf{U}^\top\mathbf{AU}$. $(\mathbf{AU})_{:1} = \mathbf{AU}_{:1} = \mathbf{Av} = \lambda\mathbf{v}$. Thus, we have $(\mathbf{U}^\top\mathbf{AU})_{:1} = \lambda\mathbf{U}^\top\mathbf{v} = \lambda\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}$ Since $\mathbf{U}$ is $\mathbf{A}$-invariant, for each i, we have $\mathbf{Au_i} \in \mathcal{U}$, which means that $\mathbf{Au_i} = \mathbf{U}\begin{pmatrix} 0 \\ \mathbf{b}_i \end{pmatrix}$ for some $\mathbf{b}_i \in \mathbb{R}^{m-1}$ (indeed, $\mathbf{Au}_i$ is orthogonal to $\mathbf{v}$). We also have $\mathbf{Av} = \lambda\mathbf{v} = \mathbf{U}\begin{pmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, from which $\mathbf{AU} = \mathbf{U}\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{pmatrix}$ follows.

(iii) Now, we want to prove that $\mathbf{B}$ is symmetric or $\mathbf{B} = \mathbf{B}^\top$. Since $\mathbf{U}$ is square and $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{UU}^\top = \mathbf{I}$. We have $\mathbf{A} = \mathbf{U}\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{pmatrix}\mathbf{U}^\top$. Then, $\mathbf{A}^\top = (\mathbf{U}\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{pmatrix}\mathbf{U}^\top)^\top$. Because we know $\mathbf{A}$ is

symmetric, $\mathbf{A}^\top = \mathbf{A} = \mathbf{U}\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B}^\top & \\ 0 & & & \end{pmatrix}\mathbf{U}^\top = \mathbf{U}\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{pmatrix}\mathbf{U}^\top$. Therefore, $\mathbf{B} = \mathbf{B}^\top$.

(iv) Since $\mathbf{B}$ is symmetric, we can reiterate the process $m-1$ times to diagonalize $\mathbf{B}$ and end up with $\mathbf{A} = \begin{pmatrix} | & & | \\ \mathbf{v_1} & \cdots & \mathbf{v_m} \\ | & & | \end{pmatrix}\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{pmatrix}\begin{pmatrix} - & \mathbf{v_1} & - \\ & \vdots & \\ - & \mathbf{v_m} & - \end{pmatrix}$ where $\mathbf{v_i}$ is the eigenvector associated with an eigenvalue λ_i for each $i \in \{1, \cdots, m\}$.

□

4 Schur decomposition and the Jordan canonical form

Definition 7 (Conjugate transpose and unitary matrices)**.** *Let* $\mathbf{U} \in \mathbb{C}^{m \times m}$*:*

- *The conjugate transpose of* $\mathbf{U}$*, denoted* $\mathbf{U}^*$*, is the transpose of the conjugate matrix* $\overline{\mathbf{U}}$*. Therefore,* $\mathbf{U}^* = \overline{\mathbf{U}}^T$
- $\mathbf{U}$ *is a unitary matrix if* $\mathbf{U}^*\mathbf{U} = \mathbf{I}$

Theorem 8 (Schur decomposition)**.** *For any matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$:*

- *There exists a unitary matrix $\mathbf{U} \in \mathbb{C}^{m\times m}$ such that $\mathbf{U}^*\mathbf{A}\mathbf{U}$ is upper triangular.*
- *There exists an orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{m\times m}$ such that:*

$$\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \begin{bmatrix} \mathbf{R_{1,1}} & \cdots & \cdots & \mathbf{R_{1,k}} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{R_{k,k}} \end{bmatrix} \tag{10}$$

where for all $i \in \{1, 2, \cdots, k\}$, $\mathbf{R_{i,i}}$ is either a 1×1 matrix (corresponding to a real eigenvalue of $\mathbf{A}$) is a 2×2 matrix (corresponding to a pair of conjugate complex eigenvalues, with the block having the form $\begin{pmatrix} \alpha & -\beta \\ \beta & -\alpha \end{pmatrix}$.

Theorem 9 (Jordan canonical form)**.** *For any matrix $\mathbf{A} \in \mathbb{C}^{m\times m}$:*

- *There is an invertible matrix $\mathbf{P} \in \mathbb{C}^{m\times m}$ such that:*

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{bmatrix} \mathbf{J}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{J}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{J}_p \end{bmatrix} \tag{11}$$

where $\mathbf{J}_i$ for all $i \in \{1, 2, \cdots, p\}$ are called Jordan blocks.

- *A Jordan block, $\mathbf{J}_i$, corresponding to eigenvalue λ_i contains repeated values of λ_i in its diagonal, and 1s above this diagonal. All other elements of the block are 0s. Visually, the Jordan blocks are of the form*

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 & 0 \\ 0 & \lambda_i & 1 & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \lambda_i \end{bmatrix} \in \mathbb{C}^{m_i \times m_i} \tag{12}$$

where λ_i corresponds to an eigenvalue of $\mathbf{A}$.

- *The eigenvalues λ_i associated with each $\mathbf{J}_i$ do not need to be distinct*
- *The number of Jordan blocks associated with an eigenvalue λ is the geometric multiplicity of λ in $\mathbf{A}$.*
- *The above decomposition is unique up to the ordering of the Jordan blocks*
- *The sum of the the sizes of each Jordan block sums to the size of $\mathbf{P}$ and $\mathbf{A}$, i.e., $m_1 + m_2 + \cdots + m_p = m$*