IFT 6760A - Lecture 4
Linear Algebra Refresher

Scribe(s): Arthur Dehgan, Adrien Mainka **Instructor:** Guillaume Rabusseau

1 Summary

In the previous lecture, we have introduced eigenvalues and eigenvectors and how they relate to matrix diagonalization. In today's lecture we will go more in depth on eigenvalues and eigenvectors and introduce the Spectral Theorem and the Schur Decomposition. Finally we will be able to prove the singular value decomposition.

2 Eigenvalues and eigenvectors (continued)

Definition 1 (Diagonalizability)**.** *A matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$ is* diagonalizable *over $\mathbb{R}$ if it is similar to a diagonal matrix :*

- *There exists an invertible matrix $\mathbf{V} \in \mathbb{R}^{m\times m}$, such that $\mathbf{V}^{-1}\mathbf{AV} = \mathbf{D}$ where D is a diagonal matrix.*
- *Or equivalently, there exists a basis $\mathbf{v_1}, ..., \mathbf{v_m}$ of $\mathbb{R}^m$ consisting of eigenvectors of $\mathbf{A}$*

Proof. We show that the two definitions are equivalent :

Let $\lambda_1, ..., \lambda_m$ the eigenvalues associated with $\mathbf{v_1}, ..., \mathbf{v_m}$ and $\mathbf{V} = \begin{bmatrix} | & & | \\ \mathbf{v_1} & \cdots & \mathbf{v_m} \\ | & & | \end{bmatrix}$ then we have :

$$\mathbf{AV} = \begin{bmatrix} | & & | \\ \lambda_1\mathbf{v_1} & \cdots & \lambda_m\mathbf{v_m} \\ | & & | \end{bmatrix} = \mathbf{VD} \tag{1}$$

Since $\mathbf{V}$ forms a basis of $\mathbb{R}^n$ it is invertible. Hence $\mathbf{V}^{-1}\mathbf{AV} = \mathbf{D}$. where $\mathbf{D}$ is a diagonal with $\mathbf{D_{i,i}} = \lambda_i$

□

Example 2.

$\mathbf{A_1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ *is diagonalizable over $\mathbb{C}$ but not over $\mathbb{R}$.*

$\mathbf{A_2} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ *is not diagonalizable. Since $\lambda = 2$, $E_\lambda = \mathcal{N}(\mathbf{A_2} - 2\mathbf{I}) = \mathcal{N}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = span\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.*

Thus we have $dim(E_\lambda) = 1$ and since $\lambda = 2$ is the only eigenvalue of $\mathbf{A_2}$ we cannot find a basis consisting of eigenvectors of $\mathbf{A}$.

Definition 3 (Eigenvalues multiplicities)**.** *Let λ be an eigenvalue of $\mathbf{A}$ then :*

- *The* geometric multiplicity, *$m_g(\lambda)$ is the dimension of E_λ*
- *The* algebric multiplicity *$m_a(\lambda)$, is the multiplicity of λ as root of $\det(\mathbf{A} - \lambda\mathbf{I})$. Equivalently, $m_a(\lambda)$ is the dimension of the generalized eigenspace $\bigcup_{k\geq 0} \mathcal{N}((\mathbf{A} - \lambda\mathbf{I})^k)$*

Example 4. *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* $m_a(\lambda) = 2$*.*

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

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 5. *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 :

$$\begin{aligned} \mathbf{A}(a\mathbf{v_1} + b\mathbf{v_2}) &= 0 \\ a\mathbf{A}\mathbf{v_1} + b\mathbf{A}\mathbf{v_2} &= 0 \\ a\lambda_1\mathbf{v_1} + b\lambda_2\mathbf{v_2} &= 0 \end{aligned} \tag{3}$$

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

$$a\lambda_2\mathbf{v_1} + b\lambda_2\mathbf{v_2} = 0 \tag{4}$$

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

$$\begin{aligned} (a\lambda_1\mathbf{v_1} + b\lambda_2\mathbf{v_2}) - (a\lambda_2\mathbf{v_1} + b\lambda_2\mathbf{v_2}) &= 0 \\ a(\lambda_1 - \lambda_2) &= 0 \end{aligned} \tag{5}$$

If the eigenvalues λ_1 and λ_2 are distinct, i.e, $\lambda_1 \neq \lambda_2$, we can conclude $a = 0$. We would use the same method to prove $b = 0$.
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* eigenvals of $\mathbf{A}$ are linearly independent. □

Definition 6 (Positive definite and semi-definite matrices)**.** *A symmetric matrix* $\mathbf{A} \in \mathbb{R}^{m\times m}$ *is :*

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

Property 7. *For any matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$*,* $\mathbf{A}^\top\mathbf{A}$ *is positive semi-definite*

Theorem 8 (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 i.e, there exists* $\mathbf{U} \in \mathbb{R}^{m\times m}$ *orthogonal such that* $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ *with* $\mathbf{D} \in \mathbb{R}^{m\times m}$ *diagonal*

Proof.

(i) Let $\mathbf{v} \neq 0$ be an eigenvector of $\mathbf{A}$, we note $\overline{\mathbf{v}}$ for the complex conjugate of $\mathbf{v}$.

$$\mathbf{v}^\top \mathbf{A}\overline{\mathbf{v}} = \mathbf{v}^\top \mathbf{A}^\top \overline{\mathbf{v}} = (\mathbf{A}\mathbf{v})^\top \overline{\mathbf{v}} = \lambda \mathbf{v}^\top \overline{\mathbf{v}} \tag{6}$$

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

$$\mathbf{v}^\top \mathbf{A}\overline{\mathbf{v}} = \mathbf{v}^\top \cdot \overline{\mathbf{A}} \cdot \overline{\mathbf{v}} = \mathbf{v}^\top \overline{\mathbf{A}\mathbf{v}} = \mathbf{v}^\top \overline{\lambda \mathbf{v}} = \overline{\lambda} \mathbf{v}^\top \overline{\mathbf{v}} \tag{7}$$

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. $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, and $||\mathbf{v}|| = 1$. Let $\mathcal{U} = span(\mathbf{v})^\perp = \{\mathbf{u} \in \mathbb{R}^m | \langle \mathbf{u}, \mathbf{v} \rangle = 0\}$. We can observe that $\mathcal{U}$ is $\mathbf{A}$ invariant, i.e. if $\mathbf{u} \in \mathcal{U}$, then $\mathbf{A}\mathbf{u} \in \mathcal{U}$. $\langle \mathbf{A}\mathbf{u}, \mathbf{v} \rangle = (\mathbf{A}\mathbf{u})^\top \mathbf{v} = \mathbf{u}^\top \mathbf{A}^\top \mathbf{v} = \mathbf{u}^\top \mathbf{A}\mathbf{v} = \lambda \mathbf{u}^\top \mathbf{v} = 0$ that implies that we have $\mathbf{A}\mathbf{u} \in \mathcal{U}$. Let $\mathbf{u_1}, ..., \mathbf{u_{m-1}} \in \mathbb{R}^m$ be an orthonormal basis of $\mathcal{U}$, and let $\mathbf{U} = \begin{pmatrix} | & | & & | \\ \mathbf{v} & \mathbf{u_1} & \dots & \mathbf{u_{m-1}} \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{m \times m}$. We notice that $\mathbf{U}$ is orthogonal, and $\mathbb{R}^m = \mathcal{U} \oplus \mathcal{U}^\top = \mathcal{U} \oplus span(\mathbf{v})$, with $dim(span(\mathbf{v})) = 1$. For each i, we have $\mathbf{A}\mathbf{u_i} \in \mathcal{U}$, which means that $\mathbf{A}\mathbf{u_i} = \mathbf{U} \begin{pmatrix} 0 \\ \mathbf{b}_i \end{pmatrix}$ for some $\mathbf{b}_i \in \mathbb{R}^{m-1}$ (indeed, $\mathbf{A}\mathbf{u}_i$ is orthogonal to $\mathbf{v}$). We also have $\mathbf{A}\mathbf{v} = \lambda\mathbf{v} = \mathbf{U} \begin{pmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, from which $\mathbf{A}\mathbf{U} = \mathbf{U} \begin{pmatrix} \lambda & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{pmatrix}$ follows. Since $\mathbf{B}$ is symmetric, we can reiterate the process $m-1$ times to diagonalize $\mathbf{B}$ and end up with $\mathbf{A} = \begin{pmatrix} | & & | \\ \mathbf{v_1} & \dots & \mathbf{v_m} \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_m \end{pmatrix} \begin{pmatrix} - & \mathbf{v_1} & - \\ & \vdots & \\ - & \mathbf{v_m} & - \end{pmatrix}$ where $\mathbf{v_i}$ is the eigenvector associated with an eigenvalue λ_i for each $i \in \{1, \dots, m\}$

□

Theorem 9 (Schur Decomposition)**.** *Let $\mathbf{A} \in \mathbb{R}^{m \times m}$. The following hold:*

- *There exists an unitary matrix $\mathbf{U} \in \mathbb{C}^{m \times m}$ such that $\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{T}$ with $\mathbf{T}$ an upper triangular matrix. Here we have $\mathbf{U}^* = \overline{\mathbf{U}}^\top$, and unitary means $\mathbf{U}^*\mathbf{U} = \mathbf{I}$.*

- *There exists a* real *orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{m \times m}$ such that :*

$$\mathbf{Q}^\top \mathbf{A}\mathbf{Q} = \begin{pmatrix} \mathbf{R_{1,1}} & \dots & \dots & \mathbf{R_{1,k}} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \mathbf{R_{k,k}} \end{pmatrix} \tag{8}$$

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

3 The Singular Value Decomposition

In a previous lecture we defined the Singular Value Decomposition (SVD), now that we have defined the spectral theorem we can prove the SVD decomposition.

3.1 Singular Value Decomposition

Theorem 10 (SVD). *Any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be written as:*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are both orthogonal matrices (i.e. $\mathbf{U}^\top\mathbf{U} = \mathbf{I_m}$ and $\mathbf{V}^\top\mathbf{V} = \mathbf{I_n}$) and $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is a diagonal rectangular matrix such that $\mathbf{\Sigma}_{i,i} \neq 0$ if and only if $i \leq rank(\mathbf{A})$

The form $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ is called the Singular-Value Decomposition (SVD) of $\mathbf{A}$.

Proof. $\mathbf{A}^\top\mathbf{A}$ is positive semi-definite and symmetric. Using the spectral theorem we can write $\mathbf{A}^\top\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^\top$ with $\mathbf{V}$ orthogonal and $\mathbf{\Sigma}$ diagonal. Let $\mathbf{A}^\top\mathbf{A} = [\mathbf{V_1}\mathbf{V_2}] \begin{pmatrix} \tilde{\mathbf{\Sigma}}^2 & \\ & \mathbf{0} \end{pmatrix} \begin{bmatrix} \mathbf{V_1}^\top \\ \mathbf{V_2}^\top \end{bmatrix}$ with $\tilde{\mathbf{\Sigma}} \in \mathbb{R}^{r \times r}$ with strictly positive entries, where r is the rank of $\mathbf{A}^\top\mathbf{A}$. It follows that $\mathbf{A}^\top\mathbf{A} = \mathbf{V_1}\tilde{\mathbf{\Sigma}}^2{\mathbf{V_1}}^\top$. Let $\mathbf{U_1} = \mathbf{A}\mathbf{V_1}\tilde{\mathbf{\Sigma}}^{-1}$. We have to prove:

(i) ${\mathbf{U_1}}^\top\mathbf{U_1} = \mathbf{I}$

(ii) ${\mathbf{U_1}}^\top\mathbf{A}\mathbf{V_1} = \tilde{\mathbf{\Sigma}}$

For (i), we have

$$\begin{aligned} {\mathbf{U_1}}^\top\mathbf{U_1} &= \tilde{\mathbf{\Sigma}}^{-1}{\mathbf{V_1}}^\top\mathbf{A}^\top\mathbf{A}\mathbf{V_1}\tilde{\mathbf{\Sigma}}^1 \\ &= \tilde{\mathbf{\Sigma}}^{-1}{\mathbf{V_1}}^\top\mathbf{V_1}\tilde{\mathbf{\Sigma}}^2{\mathbf{V_1}}^\top\mathbf{V_1}\tilde{\mathbf{\Sigma}}^{-1} \\ &= \mathbf{I} \end{aligned} \tag{9}$$

(ii) follows from the equality ${\mathbf{U_1}}^\top\mathbf{A}\mathbf{V_1} = \tilde{\mathbf{\Sigma}}^{-1}{\mathbf{V_1}}^\top\mathbf{A}^\top\mathbf{A}\mathbf{V_1} = \tilde{\mathbf{\Sigma}}$

Now we define $\mathbf{U_2} \in \mathbb{R}^{m \times (m-r)}$ orthogonal such that $\mathcal{R}(\mathbf{U_2}) = \mathcal{R}(\mathbf{U_1})^\perp$ and let $\mathbf{U} = (\mathbf{U_1}\mathbf{U_2}) \in \mathbb{R}^{m \times m}$. Observe that $\mathbf{U}$ is orthogonal by construction. We have

$$\begin{aligned} \mathbf{U}^\top\mathbf{A}\mathbf{V} &= \begin{bmatrix} {\mathbf{U_1}}^\top \\ {\mathbf{U_2}}^\top \end{bmatrix} \mathbf{A} \begin{bmatrix} {\mathbf{V_1}}^\top {\mathbf{V_2}}^\top \end{bmatrix} \\ &= \begin{bmatrix} {\mathbf{U_1}}^\top\mathbf{A}\mathbf{V_1} & {\mathbf{U_1}}^\top\mathbf{A}\mathbf{V_2} \\ {\mathbf{U_2}}^\top\mathbf{A}\mathbf{V_1} & {\mathbf{U_2}}^\top\mathbf{A}\mathbf{V_2} \end{bmatrix} \end{aligned} \tag{10}$$

We already now that the top left block is equal to $\tilde{\mathbf{\Sigma}}$. We also have

$$\mathbf{A}^\top\mathbf{A}\mathbf{V_2} = \mathbf{V_1}\tilde{\mathbf{\Sigma}}{\mathbf{v_1}}^\top = 0$$

which implies ${\mathbf{V_2}}^\top\mathbf{A}^\top\mathbf{A}\mathbf{V_2} = ||\mathbf{A}\mathbf{V_2}||_F^2 = 0$, hence $\mathbf{A}\mathbf{V_2} = 0$ and the two blocks in the second column are also equal to zero. It remains to show that ${\mathbf{U_2}}^\top\mathbf{A}\mathbf{V_1} = 0$, which follows from the fact that ${\mathbf{U_2}}^\top\mathbf{U_1} = 0 = {\mathbf{U_2}}^\top\mathbf{A}\mathbf{V_1}\tilde{\mathbf{\Sigma}}^{-1}$ since $\tilde{\mathbf{\Sigma}}^{-1}$ is an invertible matrix.
In conclusion, we have

$$\mathbf{U}^\top\mathbf{A}\mathbf{V} = \begin{bmatrix} \tilde{\mathbf{\Sigma}} & 0 \\ 0 & 0 \end{bmatrix} \tag{11}$$

□