

IFT 6760A - Lecture 5

SVD, Matrix Norms and Low Rank Approximation Theorem

Scribe(s): Joss Rakotobe, Haoyu Zhang
(adapted from Arthur Dehgan, Adrien Mainka, Erik-Olivier Riendeau
and William Dugua last year's notes)

Instructor: Guillaume Rabusseau

1 Summary

In the previous lecture, we continued reviewing the fundamentals of linear algebra. We started by covering diagonalizability. Then, we looked at the difference between the geometric and algebraic multiplicity. In particular, for a matrix to be diagonalizable, both these multiplicities must be equal. Also, we defined the definiteness property of a matrix and gave the Spectral theorem. Moreover, we covered the Shur decomposition.

In this lecture, we started by proving the singular-value decomposition. Then, we reviewed matrix norms, in particular, the p -norm and the Frobenius norm, to formalize the concept of low-rank approximation of a matrix, specifically the Ecart-Young-Mirsky theorem. Finally, we proved a Rayleigh-Ritz theorem.

2 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.

2.1 Singular Value Decomposition

Theorem 1 (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 $\Sigma_{i,i} \neq 0$ if and only if $i \leq \text{rank}(\mathbf{A})$

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

Proof. Recall that $\mathbf{A}^\top \mathbf{A}$ 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:

$$\begin{aligned}\mathbf{U}_1^\top \mathbf{U}_1 &= \tilde{\Sigma}^{-1} \mathbf{V}_1^\top \mathbf{A}^\top \mathbf{A} \mathbf{V}_1 \tilde{\Sigma}^1 \\ &= \tilde{\Sigma}^{-1} \mathbf{V}_1^\top \mathbf{V}_1 \tilde{\Sigma}^2 \mathbf{V}_1^\top \mathbf{V}_1 \tilde{\Sigma}^{-1} \\ &= \mathbf{I}\end{aligned}$$

(ii) simply follows from the equality $\mathbf{U}_1^\top \mathbf{A} \mathbf{V}_1 = \tilde{\Sigma}^{-1} \mathbf{V}_1^\top \mathbf{A}^\top \mathbf{A} \mathbf{V}_1 = \tilde{\Sigma}^{-1} \mathbf{V}_1^\top \mathbf{V}_1 \tilde{\Sigma}^2 \mathbf{V}_1^\top \mathbf{V}_1 = \tilde{\Sigma}$.

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

$$\begin{aligned}\mathbf{U}^\top \mathbf{A} \mathbf{V} &= \begin{bmatrix} \mathbf{U}_1^\top \\ \mathbf{U}_2^\top \end{bmatrix} \mathbf{A} [\mathbf{V}_1 \mathbf{V}_2] \\ &= \begin{bmatrix} \mathbf{U}_1^\top \mathbf{A} \mathbf{V}_1 & \mathbf{U}_1^\top \mathbf{A} \mathbf{V}_2 \\ \mathbf{U}_2^\top \mathbf{A} \mathbf{V}_1 & \mathbf{U}_2^\top \mathbf{A} \mathbf{V}_2 \end{bmatrix}\end{aligned}$$

By (ii), we already know that the top left block is equal to $\tilde{\Sigma}$. Since \mathbf{V} is orthogonal, for any column \mathbf{v}_2 of \mathbf{V}_2 , $\mathbf{V}_1^\top \mathbf{v}_2 = \mathbf{0}$ and

$$\mathbf{A}^\top \mathbf{A} \mathbf{v}_2 = \mathbf{V}_1 \tilde{\Sigma}^2 \mathbf{V}_1^\top \mathbf{v}_2 = \mathbf{0}$$

Hence, $\mathbf{v}_2^\top \mathbf{A}^\top \mathbf{A} \mathbf{v}_2 = \|\mathbf{A} \mathbf{v}_2\|_2^2 = 0$, which is equivalent to $\mathbf{A} \mathbf{v}_2 = \mathbf{0}$. Therefore, the two blocks in the second column are both zeros. It remains to show that $\mathbf{U}_2^\top \mathbf{A} \mathbf{V}_1 = \mathbf{0}$, which follows from the fact that $\mathbf{U}_2^\top \mathbf{U}_1 = \mathbf{0} = \mathbf{U}_2^\top \mathbf{A} \mathbf{V}_1 \tilde{\Sigma}^{-1}$ and since $\tilde{\Sigma}^{-1}$ is an invertible matrix.

In conclusion, we have

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

□

The columns of \mathbf{V} (respectively \mathbf{U}) contain the eigenvectors of $\mathbf{A}^\top \mathbf{A}$ (respectively $\mathbf{A} \mathbf{A}^\top$), also called the left (respectively right) singular vectors.

In the case where \mathbf{A} is symmetric, its SVD can be retrieved from its spectral decomposition: $\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^\top$. Since we want a diagonal matrix Σ with non negative elements (the singular values), we can define $\Sigma = |\mathbf{D}|$, where the absolute value is taken element-wise and we construct $\mathbf{V}^\top = \text{sign}(\Sigma) \mathbf{U}^\top$, where the sign function is taken element-wise.

3 Matrix Norms

3.1 Matrix p-norm

We start by defining the basic building block of the matrix p -norm:

Definition 2 (Vector p -norm). The vector p -norm, where $p \in \mathbb{R}$ is greater than 1, is defined as:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |\mathbf{x}_i|^p \right)^{1/p}.$$

See [1] for more information on vector norms.

Now, any norm on vectors induces a norm on matrices. The matrix p -norm of an arbitrary matrix \mathbf{A} , denoted $\|\mathbf{A}\|_p$, is defined as:

$$\|\mathbf{A}\|_p = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A} \mathbf{x}\|_p}{\|\mathbf{x}\|_p}.$$

Remember that the difference between the supremum and the maximum is that the maximum must be an element of the set while the supremum need not to be. More specifically, if \mathcal{X} is an ordered set, and \mathcal{S} is a subset, then s_0 is the *supremum* of \mathcal{S} iff:

1. $s \leq s_0, \forall s \in \mathcal{S}$
2. if $x \in \mathcal{X}$ and $s \leq x, \forall s \in \mathcal{S}$, then $s_0 \leq x$

On the other hand, an element m is the *maximum* of \mathcal{S} iff:

1. $s \leq m, \forall s \in \mathcal{S}$
2. $m \in \mathcal{S}$

Considering that a property of a vector norm is $\|c\mathbf{x}\|_p = |c|\|\mathbf{x}\|_p$, for any scalar c , we choose c such that $\|\mathbf{x}\|_p = 1$. Therefore, the following equivalent¹ statement defines the matrix p -norm.

Definition 3 (Matrix p -norm). *The matrix p -norm, where $p \in \mathbb{R}$, is defined as:*

$$\|\mathbf{A}\|_p = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|_p$$

Some geometric intuition about the matrix p -norm can be seen in [2].

As a side note, the main difference between a norm and a distance is that one can consider the norm of only one element, while a distance needs at least two elements.

3.2 Matrix Frobenius Norm

Definition 4 (Frobenius Norm). *The Frobenius norm is the 2-norm of the vector obtained by concatenating the rows (or equivalently the columns) of the matrix \mathbf{A} :*

$$\|\mathbf{A}\|_F = \left(\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2 \right)^{1/2}$$

From the previous definition, the Frobenius norm can also be obtained by rearranging the square of the norm in the following way:

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2 = \text{Tr}(\mathbf{A}^T \mathbf{A}) = \text{Tr}(\mathbf{A} \mathbf{A}^T)$$

Where we used the following property of traces: $\text{Tr}(\mathbf{ABC}) = \text{Tr}(\mathbf{CAB}) = \text{Tr}(\mathbf{BCA})$

Property 5. *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$ and let $\mathbf{P} \in \mathbb{R}^{m \times m}$, $\mathbf{Q} \in \mathbb{R}^{n \times n}$ be two orthogonal matrices. Then, the following holds:*

- *Matrix norms induced by vector norms: $\|\mathbf{A}\mathbf{x}\|_p \leq \|\mathbf{A}\|_p \|\mathbf{x}\|_p$*
- *Orthogonal matrices preserve the Frobenius norm: $\|\mathbf{PAQ}\|_F = \|\mathbf{A}\|_F$*
- *Orthogonal matrices preserve the 2-norm: $\|\mathbf{PAQ}\|_2 = \|\mathbf{A}\|_2$*

In particular, the last two points implies that any matrix has the same 2-norm and Frobenius norm as the diagonal rectangular matrix Σ from its SVD.

Proof. 1.

$$\begin{aligned} \|\mathbf{A}\|_p &= \sup_{\mathbf{x}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \\ &\geq \frac{\|\mathbf{A}\mathbf{y}\|_p}{\|\mathbf{y}\|_p} \\ &\Rightarrow \|\mathbf{A}\|_p \|\mathbf{y}\|_p \geq \|\mathbf{A}\mathbf{y}\|_p \forall \mathbf{y} \end{aligned}$$

¹Since we restrict the optimization to unit vectors, the subset \mathcal{S} is closed and bounded. It implies that the supremum is a maximum.

2. $\|\mathbf{PAQ}\|_F = \text{Tr}(\mathbf{Q}^\top \mathbf{A}^\top \mathbf{P}^\top \mathbf{PAQ}) = \text{Tr}(\mathbf{Q}^\top \mathbf{A}^\top \mathbf{AQ}) = \text{Tr}(\mathbf{QQ}^\top \mathbf{A}^\top \mathbf{A}) = \text{Tr}(\mathbf{A}^\top \mathbf{A}) = \|\mathbf{A}\|_F$, where we used the fact that $\mathbf{QQ}^\top = \mathbf{I}_n$ since \mathbf{Q} is a square orthogonal matrix.

3.

$$\begin{aligned}
 \|\mathbf{PAQ}\|_2 &= \sup_{\mathbf{x}, \|\mathbf{x}\|_2=1} \|\mathbf{PAQx}\|_2 \\
 &= \sup_{\mathbf{x}, \|\mathbf{x}\|_2=1} (\mathbf{x}^\top \mathbf{Q}^\top \mathbf{A}^\top \mathbf{P}^\top \mathbf{PAQx})^{\frac{1}{2}} \\
 &= \|\mathbf{AQ}\|_2 && (\mathbf{P}^\top \mathbf{P} = \mathbf{I}_m) \\
 &= \sup_{\mathbf{x}: \|\mathbf{x}\|_2=1} \|\mathbf{AQx}\|_2 \\
 &= \sup_{\mathbf{y}: \|\mathbf{y}\|_2=1} \|\mathbf{Ay}\|_2 && (\|\mathbf{Qx}\|_2 = (\mathbf{x}^\top \mathbf{Q}^\top \mathbf{Qx})^{1/2} = (\mathbf{x}^\top \mathbf{x})^{1/2} = \|\mathbf{x}\|_2) \\
 &= \|\mathbf{A}\|_2
 \end{aligned}$$

□

Property 6. For any $\mathbf{A} \in \mathbb{R}^{m \times n}$, we have $\|\mathbf{A}\|_2 = \sigma_1$, where σ_1 is the largest singular value of \mathbf{A} .

Proof. Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the SVD of \mathbf{A} with the singular values in decreasing order.

$$\begin{aligned}
 \|\mathbf{A}\|_2 &= \|\mathbf{\Sigma}\|_2 && \text{(by the last point of Property 5)} \\
 &= \max_{\|\mathbf{x}\|=1} \|\mathbf{\Sigma x}\|_2 && \text{(definition of the 2-norm)}
 \end{aligned}$$

Since for any real unit vector $\mathbf{x} = [x_1, \dots, x_n]^\top$ we have,

$$\begin{aligned}
 \|\mathbf{\Sigma x}\|_2 &= \|\left[\sigma_1 x_1 \quad \dots \quad \sigma_r x_r \quad 0 \quad \dots \quad 0\right]^\top\|_2 && \text{(with } r \text{ the rank of } \mathbf{A}) \\
 &= \left(\sum_{i=1}^r (\sigma_i x_i)^2\right)^{\frac{1}{2}} \\
 &\leq \left(\sum_{i=1}^r (\sigma_1 x_i)^2\right)^{\frac{1}{2}} && \text{(since } \sigma_1 \text{ is the largest singular value)} \\
 &= \sigma_1 \left(\sum_{i=1}^r x_i^2\right)^{\frac{1}{2}} \\
 &\leq \sigma_1 \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}} \\
 &= \sigma_1 && \text{(since } \mathbf{x} \text{ has norm 1)}
 \end{aligned}$$

Taking $\mathbf{x} = [1 \quad 0 \quad \dots \quad 0]$ gives exactly σ_1 , therefore it is the maximum.

□

Let $\mathbf{V}_{k+1} = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_{k+1} \\ | & & | \end{bmatrix}$, Where $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$ are the eigenvectors associated with the top $k + 1$ singular values.

By the Rank-Nullity Theorem:

$$\dim \mathcal{N}(\mathbf{B}) = n - \text{rank}(\mathbf{B})$$

hence,

$$\dim \mathcal{N}(\mathbf{B}) + \dim \mathcal{R}(\mathbf{V}_{k+1}) > n$$

Which implies that $\mathcal{N}(\mathbf{B}) \cap \mathcal{R}(\mathbf{V}_{k+1}) \neq \{\mathbf{0}\}$. Then, by taking a unit vector $\mathbf{x} \in \mathcal{N}(\mathbf{B}) \cap \mathcal{R}(\mathbf{V}_{k+1})$, we have:

$$\begin{aligned} \|\mathbf{A} - \mathbf{B}\|_2^2 &\geq \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2^2 \\ &= \|\mathbf{A}\mathbf{x}\|_2^2 && \text{(Since } \mathbf{x} \text{ is in } \mathcal{N}(\mathbf{B}), \text{ then } \mathbf{B}\mathbf{x} = \mathbf{0}\text{)} \\ &= \|\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}\|_2^2 && \text{(SVD of } \mathbf{A}\text{.)} \\ &= \|\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}\|_2^2 && \text{(By Property 4, since } \mathbf{U} \text{ is orthogonal)} \\ &= \sum_{i=1}^r \sigma_i^2 \langle \mathbf{v}_i, \mathbf{x} \rangle^2 \\ &= \sum_{i=k+1}^r \sigma_i^2 \langle \mathbf{v}_i, \mathbf{x} \rangle^2 && (\langle \mathbf{v}_i, \mathbf{x} \rangle = 0 \text{ for } i \leq k \text{ since } \mathbf{x} \in \mathcal{R}(\mathbf{V}_{k+1})) \\ &\geq \sigma_{k+1}^2 \sum_{i=k+1}^r \langle \mathbf{v}_i, \mathbf{x} \rangle^2 && (\sigma_{k+1} \text{ is the largest singular value of } \mathbf{\Sigma}_k) \\ &= \sigma_{k+1}^2 \|\mathbf{V}_{k+1}^T \mathbf{x}\|_2^2 && \text{(Definition of 2-norm)} \\ &= \sigma_{k+1}^2 \|\mathbf{V}^T \mathbf{x}\|_2^2 && (\mathbf{V}_{k+1}^T \mathbf{x} = \mathbf{V}^T \mathbf{x} \text{ because } \mathbf{x} \in \mathcal{R}(\mathbf{V}_{k+1})) \\ &= \sigma_{k+1}^2 \|\mathbf{x}\|_2^2 && \text{(By Property 4, since } \mathbf{V} \text{ is orthogonal)} \\ &= \sigma_{k+1}^2 && (\mathbf{x} \text{ is of unit length)} \\ &= \|\mathbf{A} - \mathbf{A}_k\|_2^2 && \text{(As shown in the first part of the proof)} \end{aligned}$$

By taking the square root on each side of the inequality, we obtain that for any matrix $\mathbf{B} = \mathbf{X}\mathbf{Y}$:

$$\|\mathbf{A} - \mathbf{B}\|_2 \geq \|\mathbf{A} - \mathbf{A}_k\|_2$$

We now show the result for Frobenius norm. The proof relies on the following inequality, known as Weyl's inequality!(we let the proof of this inequality as an exercise):

$$\sigma_{i+j-1}(\mathbf{X} + \mathbf{Y}) \leq \sigma_i(\mathbf{X}) + \sigma_j(\mathbf{Y}).$$

Let $\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ be the matrix obtained from the truncated SVD of the matrix \mathbf{A} . We want to show that \mathbf{A}_k is the rank k matrix which is the closest to \mathbf{A} in Frobenius norm. Let \mathbf{B} be any rank k matrix, we want to show that

$$\|\mathbf{A} - \mathbf{A}_k\|_F \leq \|\mathbf{A} - \mathbf{B}\|_F.$$

Applying Weyl's inequality to $\mathbf{X} = \mathbf{A} - \mathbf{B}$ and $\mathbf{Y} = \mathbf{B}$ we get

$$\sigma_{i+k}(\mathbf{A}) \leq \sigma_i(\mathbf{A} - \mathbf{B}) + \sigma_{k+1}(\mathbf{B})$$

and since \mathbf{B} is of rank k we have $\sigma_{k+1}(\mathbf{B}) = 0$, hence

$$\sigma_{i+k}(\mathbf{A}) \leq \sigma_i(\mathbf{A} - \mathbf{B}).$$

Using this inequality we obtain

$$\begin{aligned}
\|\mathbf{A} - \mathbf{A}_k\|_F &= \sum_{i=k+1}^r \sigma_i(\mathbf{A})^2 \\
&= \sum_{i=1}^{r-k} \sigma_{i+k}(\mathbf{A})^2 \\
&\leq \sum_{i=1}^{r-k} \sigma_i(\mathbf{A} - \mathbf{B})^2 \\
&\leq \sum_{i=1}^{\min(m,n)} \sigma_i(\mathbf{A} - \mathbf{B})^2 \\
&= \|\mathbf{A} - \mathbf{B}\|_F.
\end{aligned}$$

□

5 Variational Characterization of Eigenvalues of Symmetric Matrices

Definition 8 (Rayleigh Quotient). Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric, then the Rayleigh Quotient is the ratio

$$\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

The quotient is independent of the scale of \mathbf{x} since the denominator is the squared norm of \mathbf{x} .

Theorem 9 (Rayleigh-Ritz). Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric. The solution to maximizing (resp. minimizing) the Rayleigh-Ritz Quotient for $\mathbf{x} \neq \mathbf{0}$ is given by the largest (resp. smallest) eigenvalue of \mathbf{A} :

$$\max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^T \mathbf{A} \mathbf{x} = \lambda_{\max}(\mathbf{A}) \quad (1)$$

$$\min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \min_{\|\mathbf{x}\|_2=1} \mathbf{x}^T \mathbf{A} \mathbf{x} = \lambda_{\min}(\mathbf{A}) \quad (2)$$

Moreover, if $\mathbf{v}_1, \dots, \mathbf{v}_k$ are the eigenvectors corresponding to the top k eigenvalues $\lambda_1, \dots, \lambda_k$ of \mathbf{A} , then

$$\max_{\substack{\|\mathbf{x}\|_2=1 \\ \mathbf{x} \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)^\perp}} \mathbf{x}^T \mathbf{A} \mathbf{x} = \lambda_{k+1}(\mathbf{A}) \quad (3)$$

where the maximum is obtained by letting $\mathbf{x} = \mathbf{v}_{k+1}$. The constraint $\mathbf{x} \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)^\perp$ means that \mathbf{x} must be orthogonal to the first k eigenvectors of \mathbf{A} . Since \mathbf{A} is assumed to be symmetric, all its eigenvectors are orthogonal.

Proof. (1) Let $\mathbf{A} = \mathbf{V} \mathbf{D} \mathbf{V}^T$ be the eigendecomposition of \mathbf{A} , where $\mathbf{v}_1, \dots, \mathbf{v}_n$ are the eigenvectors of \mathbf{A} and \mathbf{V} is an orthogonal matrix constructed as:

$$\mathbf{V} = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Let $\mathbf{x} \in \mathbb{R}^n$ be of unit norm ($\|\mathbf{x}\|_2 = 1$) and let \mathbf{y} be a linear combination of the eigenbasis of \mathbf{A} , such that $\mathbf{y} = \mathbf{V}^T \mathbf{x} \in \mathbb{R}^n$. By **Property 5**, since \mathbf{V} is orthogonal, $\|\mathbf{y}\|_2 = \|\mathbf{x}\|_2 = 1$. Then we can derive the following inequality:

$$\begin{aligned}
\mathbf{x}^T \mathbf{A} \mathbf{x} &= \mathbf{x}^T \mathbf{V} \mathbf{D} \mathbf{V}^T \mathbf{x} && \text{(Eigendecomposition of } \mathbf{A} \text{)} \\
&= \mathbf{y}^T \mathbf{D} \mathbf{y} && \text{(Definition of } \mathbf{y} \text{)} \\
&= \sum_{i=1}^n \lambda_i \mathbf{y}_i^2 && \text{(Where } \lambda_i \text{ are the diagonal elements of } \mathbf{D} \text{)} \\
&\leq \lambda_{\max}(\mathbf{A}) \sum_{i=1}^n \mathbf{y}_i^2 && (\lambda_{\max}(\mathbf{A}) \geq \lambda_i \forall i) \\
&= \lambda_{\max}(\mathbf{A}) \|\mathbf{y}\|_2^2 \\
&= \lambda_{\max}(\mathbf{A})
\end{aligned}$$

□

Proof. (2) We are using the same matrices \mathbf{A} and \mathbf{V} , vector \mathbf{x} and \mathbf{y} as in (1). Then, we can derive the following inequality:

$$\begin{aligned}
\mathbf{x}^T \mathbf{A} \mathbf{x} &= \mathbf{x}^T \mathbf{V} \mathbf{D} \mathbf{V}^T \mathbf{x} && \text{(Eigendecomposition of } \mathbf{A} \text{)} \\
&= \mathbf{y}^T \mathbf{D} \mathbf{y} && \text{(Definition of } \mathbf{y} \text{)} \\
&= \sum_{i=1}^n \lambda_i \mathbf{y}_i^2 && \text{(Where } \lambda_i \text{ are the diagonal elements of } \mathbf{D} \text{)} \\
&\geq \lambda_{\min}(\mathbf{A}) \sum_{i=1}^n \mathbf{y}_i^2 && (\lambda_{\min}(\mathbf{A}) \leq \lambda_i \forall i) \\
&= \lambda_{\min}(\mathbf{A}) \|\mathbf{y}\|_2^2 \\
&= \lambda_{\min}(\mathbf{A})
\end{aligned}$$

□

Proof. (3) We will split the matrix \mathbf{V} in two partitions such that $\mathbf{V} = [\mathbf{V}_1 \quad \mathbf{V}_2]$, where $\mathbf{V}_1 \in \mathbb{R}^{n \times k}$ represents the top k eigenvectors of \mathbf{A} and $\mathbf{V}_2 \in \mathbb{R}^{n \times n-k}$ represents the last $n - k$ eigenvectors of \mathbf{A} . If $\mathbf{x} \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)^\perp$, then \mathbf{x} is orthogonal to all vectors in \mathbf{V}_1 and, most importantly, \mathbf{x} is in the range of \mathbf{V}_2 ($\mathbf{x} \in \mathcal{R}(\mathbf{V}_2)$). Then, we can write

$$\begin{aligned}
\mathbf{x}^T \mathbf{A} \mathbf{x} &= \mathbf{x}^T \mathbf{V}_2 \mathbf{V}_2^T \mathbf{A} \mathbf{V}_2 \mathbf{V}_2^T \mathbf{x} && \text{(Replace each } \mathbf{x} \text{ by its projection onto } \mathcal{R}(\mathbf{V}_2), \text{ which} \\
& && \text{does not change } \mathbf{x} \text{ since it is initially assumed that } \mathbf{x} \in \mathcal{R}(\mathbf{V}_2) \text{)} \quad (4) \\
&= \mathbf{x}^T \mathbf{V}_2 \mathbf{V}_2^T \mathbf{V} \mathbf{D} \mathbf{V}^T \mathbf{V}_2 \mathbf{V}_2^T \mathbf{x} && \text{(Eigendecomposition of } \mathbf{A} \text{)}
\end{aligned}$$

Now, using the fact that

$$\mathbf{V}_2^T \mathbf{V} = \mathbf{V}_2^T [\mathbf{V}_1 \quad \mathbf{V}_2] = [\mathbf{V}_2^T \mathbf{V}_1 \quad \mathbf{V}_2^T \mathbf{V}_2] = [\mathbf{0} \quad \mathbf{I}]$$

it follows that

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{V} \underbrace{\begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & \lambda_{k+1} & \\ & & & & \ddots \\ & & & & & \lambda_n \end{bmatrix}}_{\mathbf{A}} \mathbf{V}^T \mathbf{x}$$

Hence,

$$\max_{\substack{\|\mathbf{x}\|_2=1 \\ \mathbf{x} \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)^\perp}} \mathbf{x}^T \mathbf{A} \mathbf{x} = \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^T \tilde{\mathbf{A}} \mathbf{x} \quad (\text{The constraint } \mathbf{x} \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)^\perp \text{ is now incorporated into the } \tilde{\mathbf{A}}.)$$

We now have the same maximization problem as in (1),
except that $\tilde{\mathbf{A}}$ replaces \mathbf{A}

$$= \lambda_{k+1}(\mathbf{A}) \quad (\text{By (1), since the largest eigenvalue of } \tilde{\mathbf{A}} \text{ is } \lambda_{k+1})$$

□

References

- [1] J. Lambers. Mat610, lecture 2, summer 2009. <http://www.math.usm.edu/lambers/mat610/sum10/lecture2.pdf>, 2009. Accessed: 2019-01-25.
- [2] J. P. Reilly. Ece lecture notes: Matrix computations for signal processing. <http://www.ece.mcmaster.ca/faculty/reilly/ece712/ch2.pdf>, 2019. Accessed: 2019-01-25.