Mathematics for Machine Learning

Exercises 4: Matrix decompositions

4.1 Compute the determinant using the Laplace expansion (using the first row) and the Sarrus rule for

$$\mathbf{A} := \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ 0 & 2 & 4 \end{bmatrix}.$$

Using the Laplace expansion:

$$\det(\mathbf{A}) = 1 \begin{vmatrix} 4 & 6 \\ 2 & 4 \end{vmatrix} - 3 \begin{vmatrix} 2 & 6 \\ 0 & 4 \end{vmatrix} + 5 \begin{vmatrix} 2 & 4 \\ 0 & 2 \end{vmatrix} = (1 \cdot 4) + (-3 \cdot 8) + (5 \cdot 4) = 0.$$

Using the Sarrus rule:

$$\begin{aligned} \det(\mathbf{A}) &= (1 \cdot 4 \cdot 4) + (2 \cdot 2 \cdot 5) + (0 \cdot 3 \cdot 6) \\ &\quad - (0 \cdot 4 \cdot 5) - (1 \cdot 2 \cdot 6) - (2 \cdot 3 \cdot 4) \\ &= 16 + 20 + 0 - 0 - 12 - 24 \\ &= 0. \end{aligned}$$

4.2 Compute the following determinant efficiently:

$$\begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 2 & -1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ -2 & 0 & 2 & -1 & 2 \\ 2 & 0 & 0 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 2 & 0 & 1 & 2 & 0 \\ 0 & -1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & -3 \end{vmatrix} = 6.$$

4.3 Compute the eigenspaces of

$$\text{a.} \quad \mathbf{A} := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad \text{b.} \quad \mathbf{B} := \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix}$$

a. Characteristic polynomial of $\mathbf{A}$:

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = (\lambda - 1)^2 = 0 \quad \Longrightarrow \quad \lambda \in \{1\}.$$

Gives:

$$E_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^2 \mid \mathbf{A}\boldsymbol{x} = \boldsymbol{x} \right\} = \operatorname{span}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right).$$

b. Characteristic polynomial of $\mathbf{B}$:

$$\begin{aligned} p_{\mathbf{B}}(\lambda) &= \det(\mathbf{B} - \lambda\mathbf{I}) \\ &= (-2-\lambda)(\lambda-1) - 4 \\ &= \lambda^2 + \lambda - 6 && \Longrightarrow \quad \lambda \in \{2, -3\}\,. \\ &= (\lambda+3)(\lambda-2) \\ &= 0 \end{aligned}$$

Gives:

$$E_2 = \operatorname{span}\left(\begin{bmatrix}1\\2\end{bmatrix}\right), \quad E_{-3} = \operatorname{span}\left(\begin{bmatrix}2\\-1\end{bmatrix}\right).$$

4.4 Compute all eigenspaces of:

$$\mathbf{A} := \begin{bmatrix} 0 & -1 & 1 & 1 \\ -1 & 1 & -2 & 3 \\ 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 \end{bmatrix}$$

Characteristic polynomial:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda-2)(\lambda-1)(\lambda+1)^2 = 0 \quad \Longrightarrow \quad \lambda \in \{2, 1, -1\}\,.$$

$$E_2 = \operatorname{span}\left(\begin{bmatrix}1\\0\\1\\1\end{bmatrix}\right), \quad E_1 = \operatorname{span}\left(\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\right), \quad E_{-1} = \operatorname{span}\left(\begin{bmatrix}0\\1\\1\\0\end{bmatrix}\right).$$

4.5 Diagonalizability of a matrix is unrelated to its invertibility. Determine for the following four matrices whether they are diagonalizable and/or invertible:

$$\mathbf{I_2} := \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} \quad \mathbf{B} := \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} \quad \mathbf{C} := \begin{bmatrix}1 & 1\\0 & 1\end{bmatrix} \quad \mathbf{D} := \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}$$

$\mathbf{I}_2$ Invertible and diagonalizable (already diagonal).

$\mathbf{B}$ Diagonalizable but not invertible (already diagonal).

$\mathbf{C}$ Invertible but not diagonalizable (not a full basis of eigenvectors).

$\mathbf{D}$ Neither invertible nor diagonalizable (not a full basis of eigenvectors).

4.6 Compute the eigenspaces of the following transformation matrices. Are they diagonalizable?

$$\mathbf{A} := \begin{bmatrix} 2 & 3 & 0 \\ 1 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{B} := \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

a. Characteristic polynomial of $\mathbf{A}$:

$$\begin{aligned} p_{\mathbf{A}}(\lambda) &= \det(\mathbf{A} - \lambda \mathbf{I}_3) \\ &= (\lambda - 1)((\lambda - 2)(\lambda - 4) - 3) \\ &= (\lambda - 1)(\lambda^2 - 6\lambda + 5) \qquad \Longrightarrow \qquad \lambda \in \{5, 1\}\,. \\ &= (\lambda - 1)^2(\lambda - 5) \\ &= 0 \end{aligned}$$

$$E_5 = \left\{ \boldsymbol{x} \in \mathbb{R}^3 \mid \mathbf{A}\boldsymbol{x} = 5\boldsymbol{x} \right\} = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right),$$

$$E_1 = \left\{ \boldsymbol{x} \in \mathbb{R}^3 \mid \mathbf{A}\boldsymbol{x} = \boldsymbol{x} \right\} = \operatorname{span}\left(\begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} \right).$$

$\mathbf{A}$ is not diagonalizable, since it lacks a full basis of eigenvectors (specifically, $\operatorname{rank}(\mathbf{A} - \mathbf{I}) = 2$, so the geometric multiplicity of $\lambda_2 = 0$ is less than its algebraic multiplicity).

b. Characteristic polynomial of $\mathbf{B}$:

$$\begin{aligned} p_{\mathbf{B}} &= \det(\mathbf{B} - \lambda \mathbf{I}_4) \\ &= (-\lambda)(-\lambda)(-\lambda)(\lambda - 1) \qquad \Longrightarrow \qquad \lambda \in \{1, 0\}\,. \\ &= \lambda^3(\lambda - 1) \\ &= 0 \end{aligned}$$

$$E_1 = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right), \qquad E_0 = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right).$$

$\mathbf{B}$ is diagonalizable, since it has a full basis of eigenvectors:

$$\mathbf{B} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

4.7 Are the following matrices diagonalizable? If yes, determine their diagonal form, and a basis with respect to which the transformation matrices are diagonal. If no, give reasons why they are not diagonalizable.

$$\mathbf{A} := \begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix} \qquad \mathbf{B} := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \mathbf{C} := \begin{bmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{bmatrix} \qquad \mathbf{D} := \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

A Characteristic polynomial, $p_{\mathbf{A}}(\lambda) = \lambda^2 - 4\lambda + 8$, has no real roots, so **A** has no (non-trivial) eigenvectors, and is therefore not diagonalizable.

B Characteristic polynomial has two real roots, $p_{\mathbf{B}}(\lambda) = \lambda^2(\lambda - 3) = 0 \implies \lambda \in \{3, 0\}$, with matching algebraic and geometric multiplicity, so **B** diagonalizes.

$$E_3 = \operatorname{span}\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right), \qquad E_0 = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}\right) = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}\right).$$

These eigenvectors are mutually orthogonal, so we can normalize them:

$$\mathbf{Q} := \begin{bmatrix} \sqrt{3}/3 & \sqrt{2}/2 & \sqrt{6}/6 \\ \sqrt{3}/3 & -\sqrt{2}/2 & \sqrt{6}/6 \\ \sqrt{3}/3 & 0 & -\sqrt{6}/3 \end{bmatrix}, \quad \mathbf{\Lambda} := \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top.$$

C Characteristic polynomial has three real roots, $p_{\mathbf{C}} = (\lambda - 4)^2(\lambda - 2)(\lambda - 1) \implies \lambda \in \{4, 2, 1\}$. The corresponding eigenspaces are:

$$E_4 = \operatorname{span}\left(\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}\right), \quad E_2 = \operatorname{span}\left(\begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}\right), \quad E_1 = \operatorname{span}\left(\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}\right).$$

Since $\lambda_1 = 4$ has algebraic multiplicity 2, but geometric multiplicity of only 1, **C** does not diagonalize (it lacks a full basis of eigenvectors).

D Characteristic polynomial has two real roots, $p_{\mathbf{D}} = (\lambda-2)^2(\lambda-1) = 0 \implies \lambda \in \{2, 1\}$. The corresponding eigenspaces are:

$$E_2 = \operatorname{span}\left(\begin{bmatrix}2\\1\\0\end{bmatrix}, \begin{bmatrix}2\\0\\1\end{bmatrix}\right), \quad E_1 = \operatorname{span}\left(\begin{bmatrix}3\\-1\\3\end{bmatrix}\right).$$

Using Gram-Schmidt, we can orthonormalize the eigenvectors:

$$\boldsymbol{e}_1 := \begin{bmatrix}2\\1\\0\end{bmatrix}, \qquad \boldsymbol{u}_1 := \boldsymbol{e}_1 \qquad \boldsymbol{q}_1 := \frac{\boldsymbol{u_1}}{\|\boldsymbol{u_1}\|},$$

$$\boldsymbol{e}_2 := \begin{bmatrix}2\\0\\1\end{bmatrix}, \qquad \boldsymbol{u}_2 := \boldsymbol{e}_2 - \frac{\boldsymbol{u}_1\boldsymbol{u}_1^\top}{\|\boldsymbol{u}_1\|^2}\boldsymbol{e}_2, \qquad \boldsymbol{q}_2 := \frac{\boldsymbol{u}_2}{\|\boldsymbol{u}_2\|},$$

$$\boldsymbol{e}_3 := \begin{bmatrix}3\\-1\\3\end{bmatrix}, \qquad \boldsymbol{u}_3 := \boldsymbol{e}_3 - \frac{\boldsymbol{u}_1\boldsymbol{u}_1^\top}{\|\boldsymbol{u}_1\|^2}\boldsymbol{e}_3 - \frac{\boldsymbol{u}_1\boldsymbol{u}_1^\top}{\|\boldsymbol{u}_1\|^2}\boldsymbol{e}_3, \qquad \boldsymbol{q}_3 := \frac{\boldsymbol{u}_3}{\|\boldsymbol{u}_3\|}.$$

Concatenating the resulting vectors, $\boldsymbol{q}_i$, into the matrix $\mathbf{Q}$, gives us:

$$\mathbf{Q} := \begin{bmatrix} 2\sqrt{5}/5 & 2\sqrt{15}/15 & -1/3 \\ \sqrt{5}/5 & -4\sqrt{5}/15 & 2/3 \\ 0 & \sqrt{5}/3 & 2/3 \end{bmatrix}, \quad \boldsymbol{\Lambda} := \begin{bmatrix}2&0&0\\0&2&0\\0&0&1\end{bmatrix}, \quad \mathbf{D} \neq \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top.$$

…aaand this does not work at all!

$\mathbf{D}$ is not a symmetric matrix. Its eigenvectors are *not* orthogonal (unlike those of $\mathbf{B}$ earlier, which only required normalizing to form an ONB). Gram-Schmidt changes the directions of the basis, so the eigenvectors are no longer eigenvectors.

Defining instead the eigenbasis matrix, $\mathbf{P}$, we have:

$$\mathbf{P} := \begin{bmatrix}2&2&3\\1&0&-1\\0&1&3\end{bmatrix}, \qquad \mathbf{D} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1} = \begin{bmatrix}5&-6&-6\\-1&4&2\\3&-6&-4\end{bmatrix}.$$

Decomp.	$\mathbf{A} =$	Conditions	Factor properties	Use case
Spectral	$\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$	$\mathbf{A}$ symmetric	$\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$, $\boldsymbol{\Lambda}$ diag.	PCA, covariance, normal matrices
Eigen	$\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$	$\mathbf{A}$ diagonalizable	$\mathbf{P}$ invertible (*not* orth.)	General diagonalization
Schur	$\mathbf{Q}\mathbf{T}\mathbf{Q}^\top$	Any square $\mathbf{A}$	$\mathbf{Q}$ orth., $\mathbf{T}$ upper tri.	Numerical stability, QR algorithm
SVD	$\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$	Any $m \times n$	$\mathbf{U}, \mathbf{V}$ orth., $\boldsymbol{\Sigma}$ diag. (≥ 0)	Data comp., low-rank approx.
QR	$\mathbf{Q}\mathbf{R}$	Any full-rank $m \times n$	$\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}$, $\mathbf{R}$ upper tri.	Least squares, solving $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$
LU	$\mathbf{L}\mathbf{U}$	Any square, non-sing. $\mathbf{A}$	$\mathbf{L}$ l-tri. (unit diag.), $\mathbf{U}$ u-tri.	Fast solves for multiple $\boldsymbol{b}$, basis updates
Cholesky	$\mathbf{L}\mathbf{L}^\top$	$\mathbf{A}$ symmetric positive definite	$\mathbf{L}$ lower tri.	Cov. matrices, Gaussian procs, optim.

Let's instead do a QR decomposition, $\mathbf{D} = \mathbf{QR}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is upper-triangular. Normalizing the $\boldsymbol{q}_i$ as we go, the squared norm denominator disappears (since by construction $\|\boldsymbol{q}_i\|^2 = 1$). Also, in the projection, we avoid an intermediate 3×3 matrix $(\boldsymbol{q}_i \boldsymbol{q}_i^\top)$, by computing the residual coefficients r_{ij} (which we store at the specified indices in $\mathbf{R}$, which is initialized with all zeros). The normalized column vectors $\boldsymbol{q}_i$ are concatenated, $\mathbf{Q} = [\boldsymbol{q}_1\ \boldsymbol{q}_2\ \boldsymbol{q}_3]$.

We compute the following in floating-point arithmetic, rather than exactly, since the normalization results in nested roots that quickly become extremely unwieldy:

$$\boldsymbol{d}_1 := \begin{bmatrix} 5 \\ -1 \\ 3 \end{bmatrix}, \qquad \begin{aligned} &\boldsymbol{u}_1 := \boldsymbol{d}_1, \\ &r_{11} := \|\boldsymbol{u}_1\|, \end{aligned} \qquad \boldsymbol{q}_1 := \frac{\boldsymbol{u}_1}{r_{11}},$$

$$\boldsymbol{d}_2 := \begin{bmatrix} -6 \\ 4 \\ -6 \end{bmatrix}, \qquad \begin{aligned} &r_{12} := \boldsymbol{q}_1^\top \boldsymbol{d}_2, \\ &\boldsymbol{u}_2 := \boldsymbol{d}_2 - r_{12}\boldsymbol{q}_1, \\ &r_{22} := \|\boldsymbol{u}_2\|, \end{aligned} \qquad \boldsymbol{q}_2 := \frac{\boldsymbol{u}_2}{r_{22}},$$

$$\boldsymbol{d}_3 := \begin{bmatrix} -6 \\ 2 \\ -4 \end{bmatrix}, \qquad \begin{aligned} &r_{13} := \boldsymbol{q}_1^\top \boldsymbol{d}_3, \quad r_{23} := \boldsymbol{q}_2^\top \boldsymbol{d}_3, \\ &\boldsymbol{u}_3 := \boldsymbol{d}_3 - r_{13}\boldsymbol{q}_1 - r_{23}\boldsymbol{q}_2, \\ &r_{33} := \|\boldsymbol{u}_3\|, \end{aligned} \qquad \boldsymbol{q}_3 := \frac{\boldsymbol{u}_3}{r_{33}}.$$

We find,

$$\mathbf{Q} \approx \begin{bmatrix} -0.8452 & -0.4359 & -0.3094 \\ 0.1690 & -0.7671 & 0.6189 \\ -0.5071 & 0.4707 & 0.7220 \end{bmatrix}, \qquad \mathbf{R} \approx \begin{bmatrix} -5.9161 & 8.7896 & 7.4374 \\ 0 & -3.2776 & -0.8020 \\ 0 & 0 & 0.2063 \end{bmatrix}.$$

Computing $\mathbf{RQ}$ is the first step of the QR algorithm, and we see the eigenvalues start to emerge on the diagonal (compare to those we calculated exactly before):

$$\mathbf{RQ} \approx \begin{bmatrix} 2.71 & -0.66 & 12.64 \\ -0.15 & 2.14 & -2.61 \\ -0.10 & 0.10 & 0.15 \end{bmatrix}.$$

Since $\mathbf{RQ} = \mathbf{Q}^\top \mathbf{DQ}$, and we know $\mathbf{Q}$ is orthogonal by construction, this result is *similar* to $\mathbf{D}$. That is, it has the same eigenvalues. This is the first iteration of the QR algorithm.

4.8 Find the singular value decomposition (SVD) of the matrix

$$\mathbf{A} := \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$$

Compute the Gram matrix in the domain space:

$$\mathbf{A}^\top \mathbf{A} = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$$

Solve the characteristic polynomial:

$$\begin{aligned} & \det(\mathbf{A}^\top \mathbf{A} - \lambda \mathbf{I}) = 0 \\ \iff & \lambda(\lambda - 9)(\lambda - 25) = 0 \\ \iff & \lambda \in \{25, 9, 0\}\,. \end{aligned}$$

Gram matrix eigenvalues are the squared singular values, σ_i, so

$$\sigma_1 = 5, \quad \sigma_2 = 3, \quad \sigma_3 = 0.$$

Normalizing the Gram matrix eigenvectors gives the right singular vectors, $\boldsymbol{v}_i$,

$$\boldsymbol{v}_1 = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \\ 0 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} \sqrt{2}/6 \\ -\sqrt{2}/6 \\ 2\sqrt{2}/3 \end{bmatrix}, \quad \boldsymbol{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ -1/3 \end{bmatrix}.$$

Given $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$, find the left singular vectors,

$$\mathbf{U} = \mathbf{A}\mathbf{V}\boldsymbol{\Sigma}^+ = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/6 & 2/3 \\ \sqrt{2}/2 & -\sqrt{2}/6 & -2/3 \\ 0 & 2\sqrt{2}/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1/5 & 0 \\ 0 & 1/3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix}$$

So the full decomposition,

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ \sqrt{2}/6 & -\sqrt{2}/6 & 2\sqrt{2}/3 \\ 2/3 & -2/3 & -1/3 \end{bmatrix}.$$

4.9 Find the SVD of

$$\mathbf{A} := \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}.$$

First compute the covariance matrix:

$$\mathbf{A}^\top \mathbf{A} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}.$$

Then find its eigendecomposition. Solving the characteristic polynomial we have:

$$\begin{aligned} p_{\mathbf{A}^\top\mathbf{A}} &= \det(\mathbf{A}^\top\mathbf{A} - \lambda \mathbf{I}) \\ &= (5-\lambda)^2 - 9 \\ &= \lambda^2 - 10\lambda + 16 \qquad \implies \qquad \lambda \in \{8, 2\}\,. \\ &= (\lambda - 8)(\lambda - 2) \\ &= 0 \end{aligned}$$

Then we have the eigenvectors:

$$E_8 = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}, \qquad E_2 = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

Normalizing gives us the right singular vectors, which we concatenate to form $\mathbf{V}$, and find the singular values by taking square roots of the covariance matrix eigenvalues:

$$\mathbf{V} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}, \qquad \mathbf{\Sigma} = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{2} \end{bmatrix}.$$

The left singular values can be obtained like so, completing the decomposition:

$$\mathbf{U} = \mathbf{A}\mathbf{V}\mathbf{\Sigma}^+ = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

4.10 Find the rank-1 approximation of

$$\mathbf{A} := \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}.$$

This is the same matrix that we decomposed in 4.9. The rank-1 approximation can be found as follows, where: σ_1 is the largest singular value; $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ are the corresponding left and right singular vectors, respectively:

$$\tilde{\mathbf{A}}_1 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^\top = \frac{1}{2}\begin{bmatrix} 5 & 5 & 0 \\ 5 & 5 & 0 \end{bmatrix}.$$

4.11 Show that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ the matrices $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ possess the same non-zero eigenvalues.

Consider the singular value decomposition of the matrix $\mathbf{A}$ (which is guaranteed to exist):

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

Rewriting the Gram matrices $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ in terms of this decomposition, we have:

$$\begin{aligned} \mathbf{A}^\top \mathbf{A} &= (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top)^\top (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top) & \qquad \mathbf{A}\mathbf{A}^\top &= (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top)(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top)^\top \\ &= \mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top & &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top \\ &= \mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^\top, & &= \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^\top. \end{aligned}$$

Since $\mathbf{U}$ and $\mathbf{V}$ are both orthogonal, it should already be obvious that these matrices have the same eigenvalues, and that they must at least be positive. Continuing, rearrange as follows:

$$\boldsymbol{\Sigma}^2 = \mathbf{U}^\top \mathbf{A}\mathbf{A}^\top \mathbf{U}.$$

Substituting:

$$\begin{aligned} \mathbf{A}^\top \mathbf{A} &= \mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^\top \\ &= \mathbf{V}\mathbf{U}^\top (\mathbf{A}\mathbf{A}^\top)\mathbf{U}\mathbf{V}^\top. \end{aligned}$$

Then let $\mathbf{R} = \mathbf{V}\mathbf{U}^\top$, and therefore $\mathbf{R}^\top = \mathbf{U}\mathbf{V}^\top$. As the product of two orthogonal matrices, $\mathbf{R}$ must also be orthogonal, and so $\mathbf{R}^{-1} = \mathbf{R}^\top$. Finally,

$$\mathbf{A}^\top \mathbf{A} = \mathbf{R}^{-1}(\mathbf{A}\mathbf{A}^\top)\mathbf{R},$$

from which we conclude that $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ are similar (they have the same eigenvalues).

That the eigenvalues are non-zero follows by considering how the two matrices are constructed, and that they must necessarily be symmetrical, which implies that they are both positive definite.

4.12 Show that for $\boldsymbol{x} \neq 0$, Theorem 4.24 holds. That is,

$$\max_{\boldsymbol{x}} \frac{\|\mathbf{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \sigma_1,$$

where σ_1 is the largest singular value of $\mathbf{A} \in \mathbb{R}^{m \times n}$.