Q5 — Dimensionality Reduction (13 points)

Part 1 (3 points) — Best Rank-1 Approximation

Given:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad A = U\Sigma V^\top \text{ (thin SVD provided)}$$

By the Eckart-Young-Mirsky theorem, the rank-1 matrix B minimizing ‖A − B‖_F is:

$$B = \sigma_1 \, u_1 \, v_1^\top$$

Computation:

$$B = \sqrt{3} \cdot \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} & 1/\sqrt{6} & 2/\sqrt{6} \end{pmatrix}$$

Each entry: $\sqrt{3} \cdot \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{6}} = \frac{\sqrt{3}}{\sqrt{12}} = \frac{1}{2}$, and $\sqrt{3} \cdot \frac{1}{\sqrt{2}} \cdot \frac{2}{\sqrt{6}} = 1.$

$$\boxed{B = \begin{pmatrix} 1/2 & 1/2 & 1 \\ 1/2 & 1/2 & 1 \end{pmatrix}}$$

Part 2 (5 points) — Moore-Penrose Pseudoinverse

Goal: Show $A^+ = V\Sigma^+U^\top$ satisfies the four Moore-Penrose conditions.

Key fact: $\Sigma\Sigma^+ = \Sigma^+\Sigma = D$, where $D = \mathrm{diag}(I_r, 0_{k-r})$. Also $U^\top U = I_k$ and $V^\top V = I_k$.

Condition 1: $AA^+A = A$

$$AA^+A = (U\Sigma V^\top)(V\Sigma^+U^\top)(U\Sigma V^\top) = U\underbrace{\Sigma\Sigma^+\Sigma}_{=\Sigma}V^\top = U\Sigma V^\top = A \quad \checkmark$$

Condition 2: $A^+AA^+ = A^+$

$$A^+AA^+ = (V\Sigma^+U^\top)(U\Sigma V^\top)(V\Sigma^+U^\top) = V\underbrace{\Sigma^+\Sigma\Sigma^+}_{=\Sigma^+}U^\top = A^+ \quad \checkmark$$

Condition 3: (AA⁺)⊤ = AA⁺

$$AA^{+} = U(\Sigma\Sigma^{+})U^{\top} = UDU^{\top}$$

Since D is diagonal (symmetric): $(UDU^{\top})^{\top} = UD^{\top}U^{\top} = UDU^{\top}$. ✓

Condition 4: (A⁺A)⊤ = A⁺A

$$A^{+}A = V(\Sigma^{+}\Sigma)V^{\top} = VDV^{\top}$$

Same argument: $(VDV^{\top})^{\top} = VDV^{\top}$. ✓

Full column rank case: A⁺ = (A⊤A)⁻¹A⊤

When A has full column rank, all singular values are nonzero, so Σ⁺ = Σ⁻¹. Then:

$$(A^{\top}A)^{-1}A^{\top} = (V\Sigma U^{\top}U\Sigma V^{\top})^{-1}V\Sigma U^{\top} = (V\Sigma^{2}V^{\top})^{-1}V\Sigma U^{\top}$$

$$= V\Sigma^{-2}V^{\top}V\Sigma U^{\top} = V\Sigma^{-2}\Sigma U^{\top} = V\Sigma^{-1}U^{\top} = A^{+} \quad \blacksquare$$

Part 3 (5 points) — Random Matrix Multiplication: E[CR] = AB

Goal: Show that E[(CR)*{ij}] = (AB)*{ij} for all 1 ⩽ i ⩽ m, 1 ⩽ j ⩽ p.

Step 1 — Express (CR)_{ij}

$$(CR)_{ij} = \sum_{t=1}^{c} C_{it}\,R_{tj} = \sum_{t=1}^{c} \frac{A_{i,i_t}}{\sqrt{c\,p_{i_t}}} \cdot \frac{B_{i_t,j}}{\sqrt{c\,p_{i_t}}} = \sum_{t=1}^{c} \frac{A_{i,i_t}\,B_{i_t,j}}{c\,p_{i_t}}$$

Step 2 — Expectation of a single term

For any sample t, since $P(i_t = k) = p_k$:

$$E\left[\frac{A_{i,i_t}\,B_{i_t,j}}{p_{i_t}}\right] = \sum_{k=1}^{n} p_k \cdot \frac{A_{ik}\,B_{kj}}{p_k} = \sum_{k=1}^{n} A_{ik}\,B_{kj} = (AB)_{ij}$$

The sampling probabilities cancel, leaving exactly the matrix product entry.

Step 3 — Sum over c independent samples

By linearity of expectation:

$$E[(CR)_{ij}] = \frac{1}{c}\sum_{t=1}^{c} E\left[\frac{A_{i,i_t}\, B_{i_t,j}}{p_{i_t}}\right] = \frac{1}{c}\cdot c\cdot (AB)_{ij} = (AB)_{ij}$$

$$\boxed{E[CR] = AB} \quad \blacksquare$$

Intuition: The scaling by $1/\sqrt{c\,p_{i_t}}$ is designed so that the sampling probability exactly cancels out in expectation, making CR an unbiased estimator of AB regardless of the choice of distribution $\{p_i\}$. The choice of p_i affects the *variance*, not the expectation — choosing $p_i \propto \|A^{(i)}\|\cdot\|B_{(i)}\|$ minimizes variance (see Drineas et al., 2006).