5

(a) Covariance matrix in PCA is given as follows

$$\boldsymbol{C} = \frac{1}{N-1}\boldsymbol{X}\boldsymbol{X}^T \text{ where } \boldsymbol{X} = \begin{bmatrix}\overline{\boldsymbol{x}}_1 & \cdots & \overline{\boldsymbol{x}}_N\end{bmatrix} \text{ and } \overline{\boldsymbol{x}} \in \mathbb{R}^d \tag{1}$$

Now, this covariance matrix is symmetric as

$$\begin{aligned}\boldsymbol{C}^T &= \left(\frac{1}{N-1}\boldsymbol{X}\boldsymbol{X}^T\right)^T \\ &= \frac{1}{N-1}(\boldsymbol{X}^T)^T\boldsymbol{X}^T \\ &= \frac{1}{N-1}\boldsymbol{X}\boldsymbol{X}^T \\ &= \boldsymbol{C}\end{aligned} \tag{2}$$

Now, to prove that $\boldsymbol{C}$ is positive semi-definite we assume the opposite and show its infeasibility. That is, assume it is not positive semi-definite then there exists a vector $\boldsymbol{y} \in \mathbb{R}^d$ such that $\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y} < 0$.

$$\begin{aligned}\text{Now, } \exists \boldsymbol{y} \in \mathbb{R}^d | \boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y} < 0 &\Leftrightarrow \boldsymbol{y}^T\frac{1}{N-1}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{y} < 0 \\ &\Leftrightarrow \frac{1}{N-1}\boldsymbol{y}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{y} < 0 \\ &\Leftrightarrow \frac{1}{N-1}(\boldsymbol{X}^T\boldsymbol{y})^T\boldsymbol{X}^T\boldsymbol{y} < 0 \\ &\Leftrightarrow \frac{1}{N-1}\|\boldsymbol{X}^T\boldsymbol{y}\| < 0 \\ &\Leftrightarrow \|\boldsymbol{X}^T\boldsymbol{y}\| < 0 \\ &\Leftrightarrow \perp\end{aligned} \tag{3}$$

a contradiction.

(b) Let $\boldsymbol{v}_1, \boldsymbol{v}_2$ be two eigenvectors of $\boldsymbol{A}$ with distinct eigenvalues λ_1, λ_2 respectively. Now, consider

$$\begin{aligned}\lambda_1\boldsymbol{v}_1^T\boldsymbol{v}_2 &= \boldsymbol{A}\boldsymbol{v}_1^T\boldsymbol{v}_2 \\ &= \boldsymbol{A}\boldsymbol{v}_2^T\boldsymbol{v}_1 \quad (\boldsymbol{v}_1^T\boldsymbol{v}_2 \text{ is a scalar, so it is equal to its transpose}) \\ &= \lambda_2\boldsymbol{v}_2^T\boldsymbol{v}_1 \\ &= \lambda_2\boldsymbol{v}_1^T\boldsymbol{v}_2 \qquad\qquad\qquad\qquad (\text{taking transpose again})\end{aligned} \tag{4}$$

This implies

$$\begin{aligned}&\lambda_1\boldsymbol{v}_1^T\boldsymbol{v}_2 - \lambda_2\boldsymbol{v}_1^T\boldsymbol{v}_2 = 0 \\ &\Leftrightarrow (\lambda_1 - \lambda_2)\boldsymbol{v}_1^T\boldsymbol{v}_2 = 0 \\ &\Leftrightarrow \boldsymbol{v}_1^T\boldsymbol{v}_2 = 0 \qquad\qquad (\text{since eigenvalues are distinct})\end{aligned} \tag{5}$$

Hence, the eigenvectors of a symmetric matrix are orthonormal.

(c) $\tilde{\boldsymbol{x}}_i = \overline{\boldsymbol{x}} + \sum_{l=1}^{k} \boldsymbol{V}_l \alpha_{il}$.

Consider,

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} \|\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i\|_2^2 &= \frac{1}{N}\sum_{i=1}^{N} \| \underbrace{\overline{\boldsymbol{x}} + \sum_{l=1}^{k} \boldsymbol{V}_l \alpha_{il}}_{\tilde{\boldsymbol{x}}_i} - \underbrace{(\overline{\boldsymbol{x}} + \boldsymbol{V}\alpha_i)}_{\boldsymbol{x}_i} \|_2^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} \| \underbrace{\overline{\boldsymbol{x}} + \sum_{l=1}^{k} \boldsymbol{V}_l \alpha_{il}}_{\tilde{\boldsymbol{x}}_i} - \underbrace{(\overline{\boldsymbol{x}} + \sum_{l=1}^{N} \boldsymbol{V}_l \alpha_{il})}_{\boldsymbol{x}_i} \|_2^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} \left\| - \sum_{l=k+1}^{N} \boldsymbol{V}_l \alpha_{il} \right\|_2^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} \left\| \sum_{l=k+1}^{N} \boldsymbol{V}_l \alpha_{il} \right\|_2^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} \left(\left(\sum_{l=k+1}^{N} \boldsymbol{V}_l \alpha_{il} \right)^T \left(\sum_{l=k+1}^{N} \boldsymbol{V}_l \alpha_{il} \right) \right) && (\|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^T \boldsymbol{x}) \\
&= \frac{1}{N}\sum_{i=1}^{N} \left(\sum_{l=k+1}^{N} \boldsymbol{V}_l^T \boldsymbol{V}_l \alpha_{il}^2 \right) && (\boldsymbol{V}_i^T \boldsymbol{V}_j = 0 \text{ for } i \neq j) \\
&= \frac{1}{N}\sum_{i=1}^{N} \sum_{l=k+1}^{N} \alpha_{il}^2 && (\boldsymbol{V}_i^T \boldsymbol{V}_i = 1)
\end{aligned} \tag{6}$$

As α_{il} are small for $l > k$, $\frac{1}{N}\sum_{i=1}^{N} \|\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i\|_2^2$ is also small.

We know that $\boldsymbol{C} = \frac{1}{N-1}\boldsymbol{X}\boldsymbol{X}^T$, and $\boldsymbol{X} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^T$ which implies $\boldsymbol{C} = \frac{1}{N-1}\boldsymbol{V}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{V}^T$.

So, the eigenvalues λ_i for $\boldsymbol{C}$ is $\frac{1}{N-1}$ times the respective singular values of $\boldsymbol{X}$.

Using this we can derive the following

$$\frac{1}{N}\sum_{i=1}^{N} \|\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i\|_2^2 = \frac{N}{N-1} \sum_{i=k+1}^{N} \lambda_i \approx \sum_{i=k+1}^{N} \lambda_i \tag{7}$$

Again, as these eigenvalues are small, the overall error is also small.

(d) Let $\boldsymbol{X} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}^T$. Now, $\boldsymbol{C}$ is a 2×2 covariance matrix with $\boldsymbol{C}_{i,j} = \text{cov}(X_i, X_j)$

Since, X_1, X_2 are uncorrelated, $\text{cov}(X_i, X_j) = 0$ when $i \neq j$ and $\text{cov}(X_i, X_i) = \text{variance}(X_i)$.

Hence $\boldsymbol{C_1} = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}$ and $\boldsymbol{C_2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

The principal components are the eigenvectors of $\boldsymbol{C}$. In this context, it means set of uncorrelated vectors that can be used to express any $\boldsymbol{X}$.

Since, X_1, X_2 are already given as uncorrelated, the principal components for both cases are the standard unit vectors e_1, e_2 which are also the eigenvectors of $\boldsymbol{C}$.