Q2) Principal Component Analysis

(a) We denote $h_j = \sum_{k=1}^{K} \| W_k^{(j)} \|^2$ i.e., the square of $L_2$ norm of the j-th row vector in $W$.

Prove that $0 \le h_j \le 1$ and $\sum_{j=1}^{D} h_j = k$

(i) Prove that $0 \le h_j \le 1$

Each column of $W_k^{(j)}$ is an orthonormal basis vector of $\in \mathbb{R}^{D \times k}$.

Once $h_j$ is a sum of $L_2$ norms, $h_j \ge 0$ must be true, negative values would be positive due to squaring in the $L_2$ norm.

Due to the $L_2$ norms of columns being always 1, due to the orthonormal nature of those columns, the $L_2$ norms will always be $\le 1$

$\boxed{\text{Therefore, } 0 \le h_j \le 1}$ QED

(ii) Prove $\sum_{j=1}^{D} h_j = k$

$W_n = U\vec{a}_n$, therefore, $W = \begin{bmatrix} \vec{u}_1^T \vec{a}_1 & \cdots & \vec{u}_1 \vec{a}_n \\ \vdots & & \vdots \\ \vec{u}_D a_1 & \cdots & u_D \vec{a}_n \end{bmatrix}$

If we take the sum of the $L_2$ norms of each column vector $W$, we will have a bunch of 1's as solutions to the $L_2$ norm of each column vector (due to them being orthonormal basis vectors). We then have k columns in $W$, we will therefore have k number of 1's sumed up to get $\sum_{j=1}^{D} h_j = k$.

QED

(b) Prove that:

$$\max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} w_k^T \Lambda w_k = \max_{\substack{h_j \\ W^TW=I_K}} \sum_{j=1}^{D} h_j \lambda_j$$

We start with $\max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} w_k^T \Lambda w_k$

The definition of $w_k = U^T \vec{a}_k$ can be used, also $w_k^T = \vec{a}_k^T U$

we substitute to get:

$$\max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} \vec{a}_k^T U \Lambda U^T \vec{a}_k$$

We can then use the identity $U \Lambda U^T = \sum_{j=1}^{D} \lambda_j \vec{u}_j \vec{u}_j^T$ to get

$$\max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} \sum_{j=1}^{D} \vec{a}_k^T \lambda_j \vec{u}_j \vec{u}_j^T \vec{a}_k \Rightarrow \max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} \sum_{j=1}^{D} \lambda_j (a_k^T \vec{u}_j)(u_j^T \vec{a}_k)^T$$

Then by the definition of $L^2$ norm we have

$$\max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} \sum_{j=1}^{D} \lambda_j \| a_k^T \vec{u}_j \|^2$$ by the definiton of $W_k$ we have:

($L_2$ norm $W_k^T$ = $L_2$ norm $W_k$)

$$\max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} \sum_{j=1}^{D} \lambda_j \| w_k^{(j)} \|^2$$ we can flip the summations because order does not matter

$$\max_{\substack{W \\ W^TW=I_K}} \sum_{j=1}^{D} \lambda_j \sum_{k=1}^{K} \| w_k^{(j)} \|^2$$ recall the definition of $h_j = \sum_{k=1}^{K} \| w_k^{(j)} \|^2$

$$\boxed{\max_{\substack{h_j \\ W^TW=I_K}} \sum_{j=1}^{D} h_j \lambda_j}$$

QED

(c) What are the optimal $h_j$ in (3)? Show that $a_K = u_K$ $(k=1,\dots,K)$ is a sol of (3)

(i) What are the optimal $h_j$ in (3)?

An optimal Solution would be setting $a_K$ to be an eigenvector (a column of $U$ that holds all eigenvectors (aka $a_K = u_K$). There will be a 1 with a bunch of zeros that will give us a sum of 1.

There will be $\lambda_j$'s that are 1 and 0. We want to choose where the $\boxed{\text{values are}}$ $\boxed{1}$.

(ii) Show that $a_K = u_K$ is a solution of 3.

Lets start with:

$$\max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} w_k^T \Delta w_k$$ Then once $w_k = U^T \bar{a}_k$ ; $w_k^T = \bar{a}_k^T U$

$$\Rightarrow \max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{n} \bar{a}_n^T U \Delta U^T \bar{a}_n$$ We then sub our $a_n = u_k$

$$\Rightarrow \max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} u_k^T U \Delta U^T \bar{u}_k$$ from part 6, we can show it's equivalent to

$$\Rightarrow \max_{\substack{W \\ W^TW=I_K}} \sum_{k=1}^{K} \sum_{j=1}^{D} \lambda_j \| u_k^T u_j \|^2$$

So $U^T \bar{u}_k$ will have one pair come out to 1, and the rest zero.

We will have ones across the diagonal for K rows, the rest will be zeros. We then have it multiplied by the diagonal matrix $\Delta$, this will give us $\lambda_j$ that from $\lambda_1 \to \lambda_K$

We will then get $\lambda$ for each multiplication when $a_K = u_K$ versus getting a partial $\lambda$ when we have just $a_K$. This $\lambda$ will be an optimal solution.