1 PCA Handout II

Let $\mathbf{w}$ be a d dimensional projection vector and the 1 dimensional projection of points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is obtained by setting

$$y_t = \mathbf{w}^\top \mathbf{x}_t$$

If our goal is to find a w such that w is unit length (i.e. $\|\mathbf{w}\|_2 = 1$) and spread or variance of the y's is maximized then show that the optimization problem we need to solve is:

$$\text{Maximize} \quad \mathbf{w}^\top \Sigma \mathbf{w} \qquad \text{subject to } \|\mathbf{w}\|_2 = 1$$

Start here: We need to find $\mathbf{w}$ s.t. $\|\mathbf{w}\|_2 = 1$ and it maximizes the spread/varinace of y's given by:

$$\text{Variance}(y_1, \ldots, y_n) = \frac{1}{n} \sum_{t=1}^{n} \left(y_t - \frac{1}{n} \sum_{t=1}^{n} y_t \right)^2$$

Assume the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_d$ are all unit length vectors, that is, $\forall i \in [d],\ \|\mathbf{w}_i\|_2^2 = \sum_{j=1}^{d} \mathbf{w}_i[j]^2 = 1$ and are such that for any $i \neq j$, $\mathbf{w}_i \perp \mathbf{w}_j$, that is: $\sum_{k=1}^{d} \mathbf{w}_i[k] \cdot \mathbf{w}_j[k] = 0$.

It is a fact from Linear algebra that any vector in d dimensions can be written as a linear combination of the orthonormal basis vectors: $\mathbf{w}_1, \ldots, \mathbf{w}_d$. Now say the vector $\mathbf{x}_t - \mu$ is written as a linear combination of the basis vectors as:

$$\mathbf{x}_t = \mu + \sum_{j=1}^{d} \mathbf{y}_t[j]\mathbf{w}_j$$

and define the vector $\hat{\mathbf{x}}_t$ by taking $\hat{\mathbf{x}}_t - \mu$ as linear combination of only first K of the basis, that is:

$$\hat{\mathbf{x}}_t = \mu + \sum_{j=1}^{K} \mathbf{y}_t[j]\mathbf{w}_j$$

Question 2: For any K, we have shown that:

$$\frac{1}{n}\sum_{t=1}^{n} \|\hat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2 = \sum_{j=K+1}^{d} \mathbf{w}_j^\top \Sigma \mathbf{w}_j$$

using this show that

$$\sum_{j=1}^{d} \mathbf{w}_j^\top \Sigma \mathbf{w}_j = \frac{1}{n}\sum_{t=1}^{n} \|\mathbf{x}_t - \mu\|_2^2$$

Question 3: Next, using the conclusion from question 2, show that finding orthogonal directions that maximize total variance of low dimensional projections is equivalent to finding basis that minimize reconstruction error. That is, finding orthonormal W that maximizes

$$\sum_{k=1}^{K} \frac{1}{n} \sum_{t=1}^{n} \left(y_t[k] - \frac{1}{n} \sum_{s=1}^{n} y_s[k] \right)^2$$

is equivalent to finding orthonormal W that minimizes

$$\frac{1}{n} \sum_{t=1}^{n} \|\hat{\mathbf{x}}_t - \mathbf{x}_t\|_2^2$$