Sea $x \in \mathbb{R}^d$ un vector aleatorio con $\mu = \mathbb{E}[x]$, se define el centrado:

$$\tilde{x} := x - \mu \Rightarrow \mathbb{E}[\tilde{x}] = 0$$

La covarianza de x (equivalente a la de $\tilde{x}$) se define como

$$\Sigma := \mathrm{Cov}[x] = \mathbb{E}[(x-\mu)(x-\mu)^T] = \mathbb{E}[\tilde{x}\tilde{x}^T]$$

Demostrar: Sea $w \in \mathbb{R}^d$ un vector fijo (no aleatorio), entonces

$$\mathrm{Var}[w^T\tilde{x}] = w^T\Sigma w$$

Considere al escalar $z := w^T\tilde{x}$, entonces

$$\mathbb{E}[z] = \mathbb{E}[w^T\tilde{x}] = w^T\mathbb{E}[\tilde{x}] = w^T\cdot 0 = 0$$

Luego, la varianza:

$$\mathrm{Var}[z] = \mathbb{E}[z^2] - (\cancelto{0}{\mathbb{E}[z]})^2 = \mathbb{E}[z^2]$$

$$\mathbb{E}[z^2] = \mathbb{E}[(w^T\tilde{x})^2] = \mathbb{E}[(w^T\tilde{x})(w^T\tilde{x})]$$

Como $w^T\tilde{x} \in \mathbb{R}$ entonces $w^T\tilde{x} = \tilde{x}^T w$, entonces

$$\mathbb{E}[z^2] = \mathrm{Var}[z] = \mathbb{E}[(w^T\tilde{x})(\tilde{x}^T w)]$$

$a^Tbb^Ta = a^T(bb^T)a$

$$\mathbb{E}[w^T(\tilde{x}\tilde{x}^T)w] = w^T\mathbb{E}[\tilde{x}\tilde{x}^T]w = w^T\Sigma w \quad \text{QED.}$$

• $$\mathrm{tr}(W^TX^TXW) = \sum_{j=1}^{k} w_j^T X^T X w_j = \sum_{j=1}^{k} \mathrm{Var}[w_j^T\tilde{x}]$$

Minimizar error de reconstrucción = maximizar varianza de las proyecciones.

Se tienen datos $x_1, x_2, \dots, x_n \in \mathbb{R}^d$. Primero centramos:

$\tilde{x}_i = x_i - \bar{x}$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ y apilamos en una matriz:

$$\mathbb{X} = \begin{pmatrix}\tilde{x}_1^T\\ \tilde{x}_2^T \\ \vdots \\ \tilde{x}_n^T\end{pmatrix} \Rightarrow \mathbb{X}\in\mathbb{R}^{n\times d}$$

Buscamos un subespacio de dimensión k con una base ortonormal $W \in \mathbb{R}^{d\times k}$ tal que

$$W^TW = I_k : W = [w_1, w_2, \dots, w_k]$$

La proyección ortogonal de $\tilde{x}_i$ sobre ese subespacio es:

$$\hat{x}_i = WW^T\tilde{x}_i$$

La idea: Aproximar cada $\tilde{x}_i$ por medio de $\hat{x}_i$ y minimizar el error cuadrático total:

$$\min_{W^TW=I_k} \sum_{i=1}^{n} \|\tilde{x}_i - WW^T\tilde{x}_i\|^2$$

En forma matricial: $\|\cdot\|_F$: Norma de Frobenius

$$\min_{W^TW=I_k}\|\mathbb{X}-\mathbb{X}WW^T\|_F^2$$

Expandiendo mediante la identidad $\|A\|_F^2 = \mathrm{tr}(A^TA)$ se llega a:

$$\|\mathbb{X}-\mathbb{X}WW^T\|_F^2 = \underbrace{\mathrm{tr}(\mathbb{X}^T\mathbb{X})}_{\text{No depende de } W} - \mathrm{tr}(W^T\mathbb{X}^T\mathbb{X}W)$$

$$\min_{W^TW=I_k}\|\mathbb{X}-\mathbb{X}WW^T\|_F^2 = \max_{W^TW=I_k} \mathrm{tr}(W^T\mathbb{X}^T\mathbb{X}W)$$

$$\max_{W\in\mathbb{R}^{d\times k}} \mathrm{tr}(W^T\Sigma W) \text{ sujeto a } W^TW = I_k$$

El Lagrangiano:

$$\mathcal{L}(W,\Lambda) = \mathrm{tr}(W^T\Sigma W) - \mathrm{tr}(\Lambda(W^TW - I))$$

con $\Lambda \in \mathbb{R}^{k\times k}$ matriz de multiplicadores.

Por medio de diferenciales, se tiene:

$$d\mathcal{L} = 2\,\mathrm{tr}((\Sigma W - W\Lambda)^T dW)$$

Igualando a 0:

$$2\,\mathrm{tr}((\Sigma W - W\Lambda)^T dW) = 0 \ \forall dW$$

$$(\Sigma W - W\Lambda)^T dW = 0 \ \forall dW$$

$$(\Sigma W - W\Lambda)^T = 0$$

$$\Sigma W - W\Lambda = 0$$

$$\Sigma W = W\Lambda$$

Que tiene como solución

$$\Lambda = \begin{pmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \\ 0 & & \lambda_k\end{pmatrix}$$ → Matriz diagonal de autovalores tal que $\lambda_1 \ge \lambda_2 \cdots \ge \lambda_k$

$$W^* = [q_1, q_2, \dots, q_k]$$

↳ Matriz de autovectores de $\Sigma = X^TX$