$\underline{X}' \sim f(\underline{\mu}, \Sigma)$ tal que $E[\underline{X}'] = \underline{\mu}$ y $D(\underline{X}') = \Sigma$. Sea $\underline{X} = \underline{X}' - \underline{\mu}$.

$E[\underline{X}] = \underline{0}$. Sea $Z_1 = \underline{c}_1^T \underline{X}$. Queremos maximizar $Var(Z_1) = \underline{c}_1^T \Sigma \underline{c}_1$. Observemos que maximizar $Var(Z_1)$ es lo mismo que maximizar $\frac{1}{2} \underline{c}_1^T \Sigma \underline{c}_1$.

Pero $\frac{\partial}{\partial \underline{c}_1} \frac{1}{2} \underline{c}_1^T \Sigma \underline{c}_1 = \Sigma \underline{c}_1$. Luego el valor de $\underline{c}_1$ que maximiza $Var(Z_1)$ es la solución del sistema $\Sigma \underline{c}_1 = \underline{0}$. Esto es el espacio nulo de la matriz Σ. (en este caso un conjunto infinito).

Observemos que el Hessiano de $\frac{1}{2} \underline{c}_1^T \Sigma \underline{c}_1$ es $\frac{\partial}{\underline{c}_1^T \underline{c}_1} \frac{1}{2} \underline{c}_1^T \Sigma \underline{c}_1 = \Sigma$ y que bajo el supuesto de que Σ es definida positiva el determinante del Hessiano es positivo.

Busquemos ahora un $\underline{c}_1$ de norma unitaria. Es decir tal que $\underline{c}_1^T \underline{c}_1 = \|\underline{c}_1\|^2 = 1$. Ahora maximizaremos

$$Var\left(\frac{\underline{c}_1^T}{\|\underline{c}_1\|} \underline{X}\right) = \frac{\underline{c}_1^T}{\|\underline{c}_1\|} \Sigma \frac{\underline{c}_1}{\|\underline{c}_1\|} = \frac{\underline{c}_1^T \Sigma \underline{c}_1}{\|\underline{c}_1\|^2} = \frac{\underline{c}_1^T \Sigma \underline{c}_1}{\underline{c}_1^T \underline{c}_1}$$

Los valores propios y vectores propios de una matriz satisfacen $A \underline{v}_i = \lambda_i \underline{v}_i$. Los λ_i son los valores propios y los $\underline{v}_i$ son los vectores propios

Observemos que si definimos la función $f(\underline{c};\Sigma):\mathbb{R}^p\to\mathbb{R}$

$f(\underline{c};\Sigma)=\dfrac{\underline{c}^T\Sigma\,\underline{c}}{\underline{c}^T\underline{c}}$ entonces $\forall\, a\in\mathbb{R}$ se tiene que

$$f(a\underline{c};\Sigma)=f(\underline{c};\Sigma):\ \frac{(a\underline{c})^T\Sigma(a\underline{c})}{(a\underline{c})^T(a\underline{c})}=\frac{\underline{c}^T\Sigma\underline{c}}{\underline{c}^T\underline{c}}.$$

Así que sin pérdida de generalidad puede asumirse que $\|\underline{c}\|=1$.

Luego maximizaremos f sujeto a $\|\underline{c}\|^2=\sum c_i^2=1$.

El lagrangiano será

$$\mathcal{L}(\underline{c},\beta)=\underline{c}^T\Sigma\,\underline{c}-\beta(\underline{c}^T\underline{c}-1)$$

$$\frac{\partial\mathcal{L}}{\partial\underline{c}}=2\Sigma\,\underline{c}-2\beta\underline{c}\quad\text{y}\quad\frac{\partial\mathcal{L}}{\partial\underline{c}}=0\iff\boxed{\Sigma\underline{c}=\beta\underline{c}}\ (*)$$

Los valores de $\underline{c}$ y β que satisfacen esta ecuación son los vectores y valores propios de Σ, digamos $\underline{d}_1,\underline{d}_2,\dots,\underline{d}_p$ y $\lambda_1,\lambda_2,\dots,\lambda_p$.

Observemos que si $\underline{d}_i$ es un vector propio de Σ entonces

$\underline{d}_i^T\Sigma\,\underline{d}_i=\underline{d}_i^T(\lambda_i\underline{d}_i)=\lambda_i\underline{d}_i^T\underline{d}_i$, asumiendo que $\|\underline{d}_i\|=1$

entonces ~~se~~ tiene que $Var(\underline{d}_i^T\underline{X})=\lambda_i$. Luego debemos escoger $\underline{d}_i$ correspondiente al máximo valor propio.

Sean $Z_i=\underline{d}_i^T\underline{X}$ y $Z_j=\underline{d}_j\underline{X}$, entonces

$$Cov(Z_i,Z_j)=Cov(\underline{d}_i^T\underline{X},\underline{d}_j^T\underline{X})$$
$$=E[(\underline{d}_i^T\underline{X})(\underline{d}_j^T\underline{X})]-\overbrace{E[\cancel{\underline{d}_i^T\underline{X}}]}^{0}E[\underline{d}_j^T\underline{X}]$$

$$= E[\underline{d}_i^T \underline{X}\,\underline{X}^T \underline{d}_j] = \underline{d}_i^T \Sigma \underline{d}_j = \underline{d}_i^T (\underline{d}_j \lambda_j)$$

$= \lambda_j \underline{d}_i^T \underline{d}_j = 0$, pues los vectores propios son ortogonales. Para verlo notamos que

$$\Sigma \underline{d}_i = \lambda_i \underline{d}_i \quad \text{y} \quad \underline{d}_j^T \Sigma \underline{d}_i = \lambda_i \underline{d}_j^T \underline{d}_i$$

$$(\Sigma \underline{d}_j)^T \underline{d}_i = \lambda_i \underline{d}_j^T \underline{d}_i$$

$$\lambda_j \underline{d}_j^T \underline{d}_i = \lambda_i \underline{d}_j^T \underline{d}_i$$

$\Rightarrow (\lambda_j - \lambda_i)(\underline{d}_j^T \underline{d}_i) = 0$ y esto es cierto para todo i, j si y solo si $\underline{d}_j^T \underline{d}_i = 0$, pues $\lambda_i \neq \lambda_j$ en general.

De esta manera definimos $\underline{Z} = \begin{bmatrix} \underline{d}_1^T \underline{X} \\ \underline{d}_2^T \underline{X} \\ \vdots \\ \underline{d}_p^T \underline{X} \end{bmatrix}$, o

$$\underline{Z} = \begin{bmatrix} \underline{d}_1^T \\ \underline{d}_2^T \\ \vdots \\ \underline{d}_p^T \end{bmatrix}_{p\times p} \underline{X}_{p\times p} = \begin{bmatrix} \underline{d}_1 & \underline{d}_2 & \cdots & \underline{d}_p \\ | & | & & | \end{bmatrix}^T \underline{X} = D^T \underline{X},$$

donde $D \Lambda D^{-1} = \Sigma$ es la descomposición espectral de Σ. Como Σ es simétrica $D^{-1} = D^T$.

$$\left\{ \Sigma^T = (D \Lambda D^{-1})^T = (D^{-1})^T \Lambda D^T = D \Lambda D^{-1} \right.$$

Así, $Cov(D^T \underline{X}) = D^T Cov(\underline{X}) D = D^T \Sigma D = D^T D \Lambda D^T D = \Lambda$.

Observemos que $E[D^T \underline{X}] = \underline{0}$ y $Cov(D^T \underline{X}) = \underline{\Lambda}$.

Observemos que $E[\underline{1}_p^T \underline{Z}] = 0$ y $Cov(\underline{1}_p^T \underline{Z}) = \underline{1}_p^T \underline{\Lambda}\, \underline{1}_p = \sum_{i=1}^{p} \lambda_i$

La varianza de la suma de los Z_i es la suma de los λ_i.

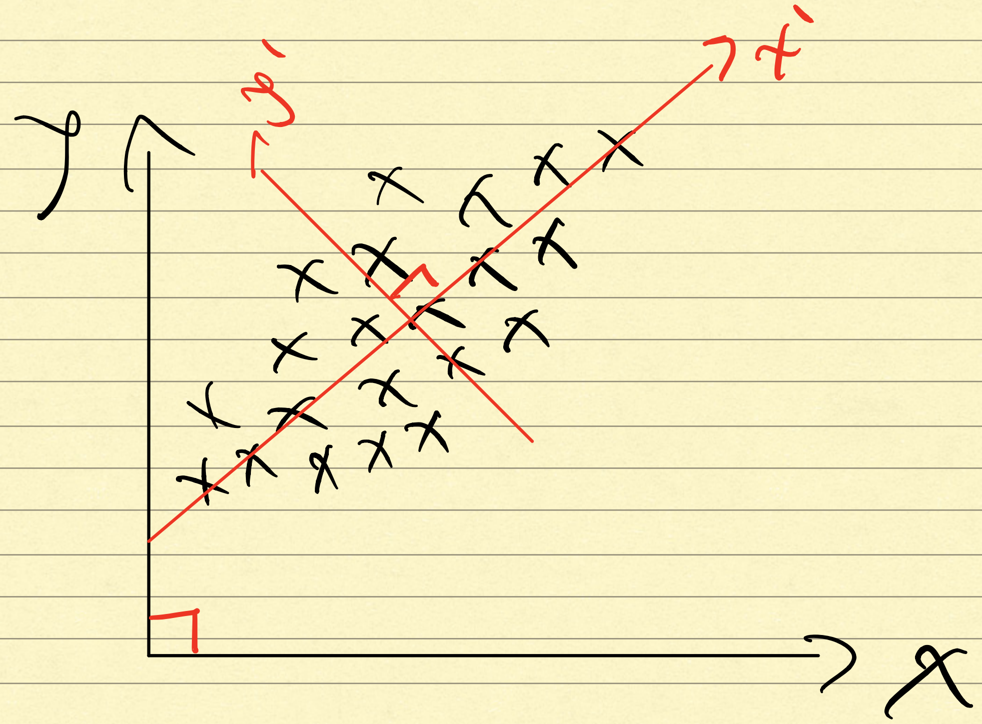
Por otro lado si $\underline{Z} = D^T \underline{X}$ entonces $\underline{X} = D\underline{Z}$.

Sea $\underline{Z}_q$ los primeros q elementos del vector $\underline{Z}$ $(q < p)$.

Así, $\underline{Z}_q = D_q^T \underline{X}$. Definimos la reconstrucción de $\underline{X}$ como

$\underline{\tilde{X}} = D_q \underline{Z}_q$. Observemos que $Cov(\underline{\tilde{X}}) = D_q\, Cov(\underline{Z}_q) D_q^T$

o sea que $Cov(\underline{\tilde{X}}) = \left(D_q \underline{\Lambda}_q D_q^T\right)_{p \times p}$



$Cov(D^T \underline{X}) = D^T \Sigma D = D^T D \underline{\Lambda} D^T D = \underline{\Lambda}$

$Cov(\underline{Z}) = \underline{\Lambda}$

$Cov(D\underline{Z}) = D\, Cov(\underline{Z}) D^T = D \underline{\Lambda} D^T = \Sigma$

Esto quiere decir que $\underline{X}$ se puede representar usando las componentes principales:

$$\underline{X} = D\,\underline{Z} = \begin{bmatrix} d_1 & d_2 & \cdots & d_p \\ | & | & & \end{bmatrix}_{p\times p} \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}_{p\times 1}$$

$$= Z_1 \cdot \underline{d_1} + Z_2 \cdot \underline{d_2} + \cdots + Z_p \cdot \underline{d_p}$$

Luego los $\underline{d_i}$ forman un diccionario para representar a los datos

Reducción de la dimensionalidad

Consideremos aproximar Σ por Σ_l donde $\Sigma_l = D_l \Lambda_l D_l^T$ donde $D_l = \begin{bmatrix} d_1 & d_2 & \cdots & d_l \\ | & | & & \end{bmatrix}$, ~~con~~

$$\Lambda_l = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_l \end{bmatrix}, \text{ con } l < p.$$

Entonces $D_l^T \underline{X} = \underline{\tilde{Z}}$ es la representación de $\underline{X}$ en el espacio de las primeras l componentes principales y $\underline{\tilde{X}} = D_l \underline{\tilde{Z}}$ es la reconstrucción de $\underline{X}$ usando las primeras l componentes principales.

¿Cuál es la calidad de la reconstrucción?

$$\| \underline{X} - \underline{\tilde{X}} \| = \| D\underline{Z} - D_l \underline{\tilde{Z}}$$