Asumiendo que $\Sigma > 0$ y tomando en cuenta que $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p$ son una base ortonormal en $\mathbb{R}^p$, pensemos en la figura A, $p=3$.

Para un dato $\mathbf{x}$ en la nube de la figura A la proyección de $\mathbf{x}$ sobre $\boldsymbol{\gamma}_i$ está dada por

$$\mathbb{P}\mathrm{roy}_{\boldsymbol{\gamma}_i}(\mathbf{x}) = (\boldsymbol{\gamma}_i' \mathbf{x}) \cdot \boldsymbol{\gamma}_i = \mathbf{y}_i \boldsymbol{\gamma}_i \qquad i=1,2,3,$$

pero además

$$(\Phi) \cdots \quad \mathbf{x} = \mathbb{P}\mathrm{roy}_{\boldsymbol{\gamma}_1}(\mathbf{x}) + \mathbb{P}\mathrm{roy}_{\boldsymbol{\gamma}_2}(\mathbf{x}) + \mathbb{P}\mathrm{roy}_{\boldsymbol{\gamma}_3}(\mathbf{x})$$

$$= \mathbf{y}_1 \boldsymbol{\gamma}_1 + \mathbf{y}_2 \boldsymbol{\gamma}_2 + \mathbf{y}_3 \boldsymbol{\gamma}_3,$$

$\mathbf{y}_i$ es el peso o el escalamiento que hay que darle al vector $\boldsymbol{\gamma}_i$, $i=1,2,3$, para obtener $\mathbf{x}$ en la ecuación (Φ), pero además $\mathbf{y}_i$ es una muestra o realización de la v.a. $\mathbb{Y}_i = \boldsymbol{\gamma}_i' \mathbb{X}$.

En particular $\mathbf{y}_1$ es una muestra de $\mathbb{Y}_1$ la primera componente principal

Para responder a la pregunta (*) en la pág 16

Supóngase ahora, que nunca vimos la Figura A, pero que tenemos los datos $x_1, \dots, x_n$. Entonces repetimos el proceso para calcular y_1, y_2 y y_3 (en la ecuación (Φ)) para cada x_i ; $i=1,2,\dots,n$. Con esto vamos a obtener una muestra de tamaño n para cada una de las componentes Y_1, Y_2 y Y_3, sean $\{y_1^i : i=1,\dots,n\}$, $\{y_2^i : i=1,\dots,n\}$ y $\{y_3^i : i=1,\dots,n\}$ estas muestras. Por construccion $VAR(Y_1) = \lambda_1 \geq \lambda_2 = VAR(Y_2) \geq \lambda_3 = VAR(Y_3)$, entonces es de esperarse que los valores $\{y_1^i\}_{i=1}^n$ se extiendan más (tengan mayor dispersión) que los valores $\{y_2^i\}_{i=1}^n$ y $\{y_3^i\}_{i=1}^n$ y que a su vez los valores $\{y_2^i\}_{i=1}^n$ tengan mayor dispersión que los valores $\{y_3^i\}_{i=1}^n$. Como $\{y_1^i\}_{i=1}^n$ son los escalamientos en la dirección de γ_1, $\{y_2^i\}_{i=1}^n$ son los escalamientos en la dirección de γ_2 y $\{y_3^i\}_{i=1}^n$ son los escalamientos en la dirección de γ''_3, entonces es de esperarse que la configuración de los datos sea como en la figura A.

NOTA: Para datos tales que $\mathbb{X} \sim F$ donde $\mathbb{E}(\mathbb{X}) \neq 0$, tendríamos que dibujar el elipsoide de la figura A en otra localización de $\mathbb{R}^3$. Lo que aparece como el origen $(0,0,0)$ tendría que convertirse en el vector $(\mathbb{E}(X_1), \mathbb{E}(X_2), \mathbb{E}(X_3))$ $p=3$.

En general los datos tendrán una distribución F para la cual la media no es el vector $\mathbb{O}$, es decir $x_1, \ldots, x_n$ son una muestra de $\mathbb{X} \sim F$ con $\mathbb{E}(\mathbb{X}) \neq 0$.

En algunos libros se asume $\mathbb{E}(\mathbb{X}) = 0$ porque de no ser así, se puede hacer teoría con el vector aleatorio centrado $\tilde{\mathbb{X}}' = \mathbb{X} - \mathbb{E}(\mathbb{X})$

y en tal caso $\mathbb{E}(\tilde{\mathbb{X}}) = \mathbb{0}$ porque

$$\mathbb{E}(\tilde{\mathbb{X}}) = \mathbb{E}(\mathbb{X} - \mathbb{E}(\mathbb{X})) = \mathbb{E}(\mathbb{X}) - \mathbb{E}(\mathbb{E}(\mathbb{X}))$$

$$= \mathbb{E}(\mathbb{X}) - \mathbb{E}(\mathbb{X}) = \mathbb{0}$$

Vamos a escribir las relaciones (PC1), (PC2) (PC3), etc... que definen las componentes principales de un vector aleatorio $\mathbb{X}$ usando Σ y μ (la matriz de varianzas-covarianzas y el vector de medias) que son características de F, la distribución de $\mathbb{X}$.

$$\mu = \mathbb{E}(\mathbb{X})$$

descomposición de Jordan

de dim. $p \times p$ simétrica y positivo def. $\longrightarrow$ $\Sigma = \mathrm{VAR}(\mathbb{X}) = \Gamma \Lambda \Gamma'$

$$\Lambda = \mathrm{DIAG}(\lambda_1, \ldots, \lambda_p)$$

$$\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$$

Las relaciones (PC1), (PC2), (PC3), ... se pueden escribir en forma matricial como

$$\mathbb{Y} = \Gamma'(\mathbb{X} - \mu)$$