MAT-269: Sesión 16, Modelo de Análisis Factorial

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Idea:

El objetivo es intentar explicar la correlación entre un conjunto grande de variables en términos de un número pequeño de factores.

Este modelo supone que aquellos factores no son observables y son considerados variables aleatorias. Este modelo es muy apropiado en áreas como Psicología.

Definición 1 (Modelo de análisis factorial)

Sea $\boldsymbol{x} = (x_1, \dots, x_p)^\top$ vector aleatorio con $\mathsf{E}(\boldsymbol{x}) = \boldsymbol{\mu}$ y $\mathsf{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$. Suponga que los elementos de $\boldsymbol{x}$ satisfacen:

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z} + \boldsymbol{\epsilon},$$

donde $\boldsymbol{z} = (z_1, \dots, z_m)^\top$ es llamado factores comunes $(m < p)$ y $\boldsymbol{\Gamma} = (\gamma_{jk}) \in \mathbb{R}^{p \times m}$, con γ_{jk} llamado la carga factorial (loadings) de la variable j (x_j) sobre el factor z_k.

El modelo está basado en los siguientes supuestos:

$$\mathsf{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \qquad \mathsf{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} = \operatorname{diag}(\psi_1^2, \dots, \psi_p^2),$$

independientes de $\boldsymbol{z}$ con

$$\mathsf{E}(\boldsymbol{z}) = \mathbf{0}, \qquad \mathsf{Cov}(\boldsymbol{z}) = \boldsymbol{I}.$$

De este modo,

$$\mathsf{E}(\boldsymbol{x}) = \boldsymbol{\mu} + \boldsymbol{\Gamma}\,\mathsf{E}(\boldsymbol{z}) + \mathsf{E}(\boldsymbol{\epsilon}) = \boldsymbol{\mu},$$

y

$$\begin{aligned}\mathsf{Cov}(\boldsymbol{x}) &= \mathsf{Cov}(\boldsymbol{\Gamma}\boldsymbol{z} + \boldsymbol{\epsilon}) = \boldsymbol{\Gamma}\,\mathsf{Cov}(\boldsymbol{z})\boldsymbol{\Gamma}^\top + \mathsf{Cov}(\epsilon)\\ &= \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \boldsymbol{\Psi} = \boldsymbol{\Sigma}.\end{aligned}$$

Ejemplo: (Spearman, 1904)

Se desea examinar el desempeño de un grupo de niños en Classics (x_1), French (x_2) y English (x_3). Se calculó la matriz

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.00 & 0.83 & 0.78 \\ 0.83 & 1.00 & 0.67 \\ 0.78 & 0.67 & 1.00 \end{pmatrix},$$

se puede verificar que $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \boldsymbol{\Sigma}$ con

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0.983 \\ 0.844 \\ 0.744 \end{pmatrix}, \qquad \boldsymbol{\Psi} = \begin{pmatrix} 0.34 & 0.00 & 0.00 \\ 0.00 & 0.29 & 0.00 \\ 0.00 & 0.00 & 0.37 \end{pmatrix},$$

luego el modelo asume la forma

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z} + \boldsymbol{\epsilon},$$

es decir

$$x_1 = \mu_1 + \gamma_1 z + \epsilon_1, \quad x_2 = \mu_2 + \gamma_2 z + \epsilon_2, \quad x_3 = \mu_3 + \gamma_3 z + \epsilon_3,$$

con z el factor "inteligencia".

Observación:

Sea $\boldsymbol{P}$ matriz ortogonal $m \times m$, entonces,

$$\begin{aligned}\boldsymbol{x} &= \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z} + \boldsymbol{\epsilon} = \boldsymbol{\mu} + (\boldsymbol{\Gamma}\boldsymbol{P})(\boldsymbol{P}^\top\boldsymbol{z}) + \boldsymbol{\epsilon} \\ &= \boldsymbol{\mu} + \boldsymbol{\Gamma}_*\boldsymbol{z}_* + \boldsymbol{\epsilon},\end{aligned}$$

donde

$$\operatorname{Cov}(\boldsymbol{z}_*) = \boldsymbol{P}^\top \operatorname{Cov}(\boldsymbol{z})\boldsymbol{P} = \boldsymbol{P}^\top\boldsymbol{P} = \boldsymbol{I},$$

de este modo el modelo no es único, es decir, existe un problema de identificabilidad.

El problema de no unicidad puede ser resuelto mediante rotar las cargas factoriales de tal manera que se satisfaga

$$\boldsymbol{\Gamma}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma} \qquad \text{sea diagonal},$$

esto incorpora $m(m-1)/2$ restricciones.

Note que los elementos diagonales de $\mathbf{\Sigma}$ $(= \mathbf{\Gamma}\mathbf{\Gamma}^\top + \mathbf{\Psi})$ asumen la forma

$$\sigma_{ii} = \sum_{k=1}^{m} \gamma_{ik}^2 + \psi_i^2 = h_i^2 + \psi_i^2,$$

es decir, podemos descomponer la varianza de x_i en dos partes h_i^2 llamada comunalidad o varianza común y ψ_i^2 es llamada varianza residual.

Estimación ML

Suponga una muestra $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ muestra aleatoria desde $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces

$$\ell(\boldsymbol{\theta}) = -\frac{np}{2} \log 2\pi - \frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}),$$

es conveniente perfilar la verosimilitud en $\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$, obteniendo

$$\ell(\boldsymbol{\Gamma}, \boldsymbol{\Psi}) = -\frac{np}{2} \log 2\pi - \frac{n}{2} \log |\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \boldsymbol{\Psi}| - \frac{1}{2} \operatorname{tr}(\boldsymbol{\Gamma}^\top \boldsymbol{\Gamma} + \boldsymbol{\Psi})^{-1} \boldsymbol{Q},$$

donde $\boldsymbol{Q} = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top$.

Se debe optimizar $\ell(\mathbf{\Gamma}, \mathbf{\Psi})$ sujeto a la condición que $\mathbf{\Gamma}^\top \mathbf{\Psi}^{-1} \mathbf{\Gamma} = \mathbf{\Delta}$, $\mathbf{\Delta}$ diagonal. Tomando derivadas con respecto a $\mathbf{\Gamma}$ y $\mathbf{\Psi}$ lleva a las siguientes ecuaciones:

$$\begin{aligned}
\boldsymbol{S}\widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{\Gamma}} &= \widehat{\mathbf{\Gamma}}(\boldsymbol{I} + \widehat{\mathbf{\Gamma}}^\top \widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{\Gamma}}) \\
\widehat{\mathbf{\Psi}} &= \operatorname{diag}(\boldsymbol{S} - \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Gamma}}^\top) \\
\widehat{\mathbf{\Delta}} &= \widehat{\mathbf{\Gamma}}^\top \widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{\Gamma}}.
\end{aligned}$$

Estas ecuaciones deben ser resueltas de forma iterativa y es bien conocido que sufre de problemas de convergencia.

Observación:

La función de log-verosimilitud puede no tener máximo, sujeto a $\mathbf{\Psi} > \mathbf{0}$, frecuentemente esto se debe a que uno o más elementos de $\mathbf{\Psi}$ tienden a cero.

Hasta que surgiera el algoritmo propuesto por Jöreskog (1967) y Lawley (1967) existía mucha dificultad en la estimación para el modelo factorial.[1]

[1] Estas rutinas se encuentran implementadas en el software LISREL.

Estimación ML: Método de componentes principales

Note que $\boldsymbol{\Sigma} - \boldsymbol{\Psi} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$ es semidefinida positiva de rango m. De ahí que existe una matriz ortogonal $\boldsymbol{M}$ tal que

$$\boldsymbol{M}^\top(\boldsymbol{\Sigma} - \boldsymbol{\Psi})\boldsymbol{M} = \operatorname{diag}(\phi_1, \dots, \phi_m, 0, \dots, 0) = \boldsymbol{\Phi},$$

donde ϕ_j son los valores propios positivos. Sea $\boldsymbol{M}_1$ la matriz que contiene los primeros m vectores propios y $\boldsymbol{\Phi}_1 = \operatorname{diag}(\phi_1, \dots, \phi_m)$. Entonces

$$\begin{aligned}\boldsymbol{\Sigma} - \boldsymbol{\Psi} = \boldsymbol{M}\boldsymbol{\Phi}\boldsymbol{M}^\top &= (\boldsymbol{M}_1, \boldsymbol{M}_2)\begin{pmatrix}\boldsymbol{\Phi}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{M}_1^\top \\ \boldsymbol{M}_2^\top\end{pmatrix} \\ &= \boldsymbol{M}_1\boldsymbol{\Phi}_1\boldsymbol{M}_1^\top = (\boldsymbol{\Gamma}_1\boldsymbol{\Phi}_1^{1/2})(\boldsymbol{\Gamma}_1\boldsymbol{\Phi}_1^{1/2})^\top,\end{aligned}$$

y $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1\boldsymbol{\Phi}_1^{1/2}$ es una solución. De este modo, para un estimador dado $\widehat{\boldsymbol{\Psi}}_1$ de $\boldsymbol{\Psi}$ se obtiene las primeras m componentes principales de $\widehat{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Psi}}_1$ lo que permite estimar $\boldsymbol{\Gamma}$ como $\widehat{\boldsymbol{\Gamma}}_1\widehat{\boldsymbol{\Phi}}_1^{1/2}$.

Considere el modelo de análisis factorial,

$$\boldsymbol{x}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z}_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \dots, n,$$

asumiremos que $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$ son vectores no observables (missing) IID $\mathsf{N}_m(\mathbf{0}, \boldsymbol{I})$, independiente de los errores $\boldsymbol{\epsilon}_i \overset{\text{IID}}{\sim} \mathsf{N}_p(\mathbf{0}, \boldsymbol{\Psi})$, con $\boldsymbol{\Psi} = \operatorname{diag}(\psi_1^2, \dots, \psi_p^2)$.

De este modo, condicional a los $\boldsymbol{z}_i$, tenemos:

$$\boldsymbol{x}_i | \boldsymbol{z}_i \overset{\text{IND}}{\sim} \mathsf{N}_p(\boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z}_i, \boldsymbol{\Psi}), \qquad \boldsymbol{z}_i \overset{\text{IND}}{\sim} \mathsf{N}_m(\mathbf{0}, \boldsymbol{I}).$$

Integrando con relación a $\boldsymbol{z}_i$ obtenemos la distribución marginal

$$\boldsymbol{x}_i \sim \mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \boldsymbol{\Psi}), \qquad i = 1, \dots, n.$$

De este modo, el vector de datos completos $(\boldsymbol{x}, \boldsymbol{z})$ con $\boldsymbol{z} = (\boldsymbol{z}_1^\top, \dots, \boldsymbol{z}_n^\top)^\top$ el vector de datos perdidos, podemos notar que

$$\boldsymbol{z}_i | \boldsymbol{x}_i \sim \mathsf{N}(\boldsymbol{\Gamma}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}), (\boldsymbol{\Gamma}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma} + \boldsymbol{I})^{-1}).$$

En este caso tenemos que la log-verosimilitud de datos completos asume la forma

$$\begin{aligned} \ell_{\mathsf{c}}(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{z}) = & -\frac{n}{2} \log |2\pi \boldsymbol{\Psi}| \\ & - \frac{1}{2} \sum_{i=1}^n \big\{ (\boldsymbol{x}_i - \boldsymbol{\mu} - \boldsymbol{\Gamma} \boldsymbol{z}_i)^\top \boldsymbol{\Psi}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu} - \boldsymbol{\Gamma} \boldsymbol{z}_i) + \boldsymbol{z}_i^\top \boldsymbol{z}_i \big\}. \end{aligned}$$

Note que,

$$\mathsf{E}(\boldsymbol{z}_i | \boldsymbol{x}_i) = (\boldsymbol{\Gamma}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma} + \boldsymbol{I})^{-1} \boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}).$$

Estimación ML: Algoritmo EM

Substituyendo $\boldsymbol{\mu}$ por $\overline{\boldsymbol{x}}$ y como:

$$
\begin{aligned}
(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\boldsymbol{z}_i)^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\boldsymbol{z}_i) &= (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \\
&\quad + \boldsymbol{z}_i^\top \boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}\boldsymbol{z}_i - 2(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}\boldsymbol{z}_i,
\end{aligned}
$$

usando resultados básicos de esperanzas de formas cuadráticas, tenemos

$$
\begin{aligned}
\mathsf{E}\{(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\boldsymbol{z}_i)^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\boldsymbol{z}_i)|\boldsymbol{x}_i, \boldsymbol{\theta}^{(k)}\} &= (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}) \\
- 2(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i + \operatorname{tr}\boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}\,\mathsf{Cov}(\boldsymbol{z}_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(k)}) &+ \widehat{\boldsymbol{z}}_i^\top \boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i \\
= (\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i)^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i) &+ \operatorname{tr}\boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}\widehat{\boldsymbol{\Omega}},
\end{aligned}
$$

donde

$$
\begin{aligned}
\widehat{\boldsymbol{z}}_i &= \mathsf{E}(\boldsymbol{z}_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(k)}) = (\widehat{\boldsymbol{\Gamma}}^\top \widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{\Gamma}} + \boldsymbol{I})^{-1}\widehat{\boldsymbol{\Gamma}}^\top \widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}), \\
\widehat{\boldsymbol{\Omega}} &= \mathsf{Cov}(\boldsymbol{z}_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(k)}) = (\widehat{\boldsymbol{\Gamma}}^\top \widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{\Gamma}} + \boldsymbol{I})^{-1},
\end{aligned}
$$

con $\widehat{\boldsymbol{\Gamma}}$ y $\widehat{\boldsymbol{\Psi}}$ siendo evaluadas en $\boldsymbol{\theta}^{(k)}$, y análogamente

$$
\mathsf{E}(\boldsymbol{z}_i^\top \boldsymbol{z}_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(k)}) = \widehat{\boldsymbol{z}}_i^\top \widehat{\boldsymbol{z}}_i + \operatorname{tr}\widehat{\boldsymbol{\Omega}}.
$$

Estimación ML: Algoritmo EM

De este modo,

$$
\begin{aligned}
Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(k)}) &= -\frac{n}{2}\log|\boldsymbol{\Psi}| \\
&\quad - \frac{1}{2}\sum_{i=1}^{n}\big\{(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i)^\top \boldsymbol{\Psi}^{-1}(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i) + \operatorname{tr}\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}.
\end{aligned}
$$

Tenemos que

$$
\begin{aligned}
\mathsf{d}_{\Psi}\, Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(k)}) &= -\frac{n}{2}\operatorname{tr}\boldsymbol{\Psi}^{-1} + \frac{n}{2}\operatorname{tr}\boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma}\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^\top\boldsymbol{\Psi}^{-1}\,\mathsf{d}\,\boldsymbol{\Psi} \\
&\quad + \frac{1}{2}\operatorname{tr}\boldsymbol{\Psi}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i)(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i)^\top\boldsymbol{\Psi}^{-1}\,\mathsf{d}\,\boldsymbol{\Psi} \\
&= \frac{1}{2}\operatorname{tr}\boldsymbol{\Psi}^{-1}\big\{\boldsymbol{S} + n\boldsymbol{\Gamma}\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^\top - n\boldsymbol{\Psi}\big\}\boldsymbol{\Psi}^{-1}\,\mathsf{d}\,\boldsymbol{\Psi}.
\end{aligned}
$$

De ahí que

$$
\boldsymbol{\Psi}^{(k+1)} = \operatorname{diag}\left(\frac{1}{n}\boldsymbol{S} + \boldsymbol{\Gamma}^{(k)}\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^{(k)\top}\right)
$$

Por otro lado,

$$
\begin{aligned}
\mathsf{d}_\Gamma\, Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(k)}) &= \sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i)^\top \boldsymbol{\Psi}^{-1}(\mathsf{d}\,\boldsymbol{\Gamma})\widehat{\boldsymbol{z}}_i \\
&\quad - \frac{n}{2}\operatorname{tr}\widehat{\boldsymbol{\Omega}}(\mathsf{d}\,\boldsymbol{\Gamma})^\top \boldsymbol{\Psi}^{-1}\boldsymbol{\Gamma} - \frac{n}{2}\operatorname{tr}\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1}\,\mathsf{d}\,\boldsymbol{\Gamma} \\
&= \sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i)^\top \boldsymbol{\Psi}^{-1}(\mathsf{d}\,\boldsymbol{\Gamma})\widehat{\boldsymbol{z}}_i - n\operatorname{tr}\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1}\,\mathsf{d}\,\boldsymbol{\Gamma},
\end{aligned}
$$

es decir,

$$
\mathsf{d}_\Gamma\, Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(k)}) = \operatorname{tr}\sum_{i=1}^n \widehat{\boldsymbol{z}}_i(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i)^\top \boldsymbol{\Psi}^{-1}\,\mathsf{d}\,\boldsymbol{\Gamma} - n\operatorname{tr}\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^\top \boldsymbol{\Psi}^{-1}\,\mathsf{d}\,\boldsymbol{\Gamma}.
$$

Desde la condición de primer orden, sigue que

$$
\sum_{i=1}^n \widehat{\boldsymbol{z}}_i(\boldsymbol{x}_i - \overline{\boldsymbol{x}} - \boldsymbol{\Gamma}\widehat{\boldsymbol{z}}_i)^\top - n\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^\top = \boldsymbol{0}
$$

$$
\sum_{i=1}^n \widehat{\boldsymbol{z}}_i(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top - \sum_{i=1}^n \widehat{\boldsymbol{z}}_i\widehat{\boldsymbol{z}}_i^\top \boldsymbol{\Gamma}^\top - n\widehat{\boldsymbol{\Omega}}\boldsymbol{\Gamma}^\top = \boldsymbol{0}
$$

Es decir,

$$n\Big(\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{z}}_i\widehat{\boldsymbol{z}}_i^\top+\widehat{\boldsymbol{\Omega}}\Big)\boldsymbol{\Gamma}^\top=\sum_{i=1}^{n}\widehat{\boldsymbol{z}}_i(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^\top.$$

Finalmente,

$$\boldsymbol{\Gamma}^{(k+1)\top}=\Big(\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{z}}_i\widehat{\boldsymbol{z}}_i^\top+\widehat{\boldsymbol{\Omega}}\Big)^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{z}}_i(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^\top.$$

Observaciones:

- Este algoritmo EM es simple y numéricamente estable.
- Posibilidad de soluciones múltiples (que no necesariamente son rotaciones unas de otras).
- Lenta razón de convergencia, lentitud que se atribuye a la proporción de datos perdidos.
- Errores estándar no son obtenidos como un subproducto del procedimiento de estimación (diferente a algoritmo Newton-Raphson).