EM algorithm for Factor Analysis

We assume centered data, i.e. $\mu = 0$, as we know that ML estimation for μ is given by the sample mean $\mu = \frac{1}{N}\sum_n x_n$.

- Complete-data log-likelihood :

$$\ln p(\underline{x}, \underline{z}; \theta) = \sum_{n=1}^{N} \left[\ln p(x_n | z_n; \theta) + \ln p(z_n) \right].$$

$$\stackrel{c}{=} \sum_{n=1}^{N} \left[\ln \mathcal{N}(x_n; W z_n, \Psi) \right].$$

we drop cst w.r.t. θ.

$$\stackrel{c}{=} -\frac{1}{2}\sum_{n=1}^{N} \left[\ln \det(\Psi) + (x_n - W z_n)^T \Psi^{-1} (x_n - W z_n) \right]$$

$$\stackrel{c}{=} -\frac{1}{2}\sum_{n=1}^{N} \left[\ln \det(\Psi) + x_n^T \Psi^{-1} x_n - z_n^T W^T \Psi^{-1} x_n - x_n^T \Psi^{-1} W z_n + z_n^T W^T \Psi^{-1} W z_n \right]$$

$$\stackrel{c}{=} -\frac{1}{2}\sum_{n=1}^{N} \left[\ln \det(\Psi) + x_n^T \Psi^{-1} x_n - 2 z_n^T W^T \Psi^{-1} x_n + tr\left(z_n z_n^T W^T \Psi^{-1} W \right) \right]$$

where we used $z_n^T W^T \Psi^{-1} x_n = (z_n^T W^T \Psi^{-1} x_n)^T = x_n^T \Psi^{-1} W z_n$ since Ψ^{-1} is diagonal.

* E-Step :

$$p(z \mid x ; \tilde{\theta}) = \prod_n p(z_n \mid x_n ; \tilde{\theta})$$

$$p(z_n \mid x_n ; \tilde{\theta}) = \mathcal{N}(z_n ; \tilde{\mu}_n , \tilde{\Sigma}_n)$$

$$\begin{cases} \tilde{\mu}_n = W^T (WW^T + \Psi)^{-1} x_n \\ \tilde{\Sigma}_n = I - W^T (WW^T + \Psi)^{-1} W \end{cases}$$

$$Q(\theta, \tilde{\theta}) = \mathbb{E}_{p(z \mid x ; \tilde{\theta})} \left[\ln p(x, z ; \theta) \right]$$

$$= -\frac{1}{2} \sum_{n=1}^{N} \left[\ln \det(\Psi) + x_n^T \Psi^{-1} x_n - 2 \mathbb{E}[z_n]^T W^T \Psi^{-1} x_n + \operatorname{tr}\left(\mathbb{E}[z_n z_n^T] W^T \Psi^{-1} W \right) \right]$$

$$\left(\cdot \ \mathbb{E}[z_n] = \tilde{\mu}_n \quad , \quad \cdot \ \mathbb{E}[z_n z_n^T] = \tilde{\mu}_n \tilde{\mu}_n^T + \tilde{\Sigma}_n \right)$$

$$= -\frac{1}{2} \sum_{n=1}^{N} \left[\ln \det(\Psi) + x_n^T \Psi^{-1} x_n - 2 \tilde{\mu}_n^T W^T \Psi^{-1} x_n + \operatorname{tr}\left([\tilde{\mu}_n \tilde{\mu}_n^T + \tilde{\Sigma}_n] W^T \Psi^{-1} W \right) \right]$$

M-Step

$$\frac{\partial Q}{\partial W} = \frac{\partial}{\partial W}\left\{ -\frac{1}{2}\sum_{n=1}^{N}\left(-2\,\mathrm{tr}\left[W^T\Psi^{-1}x_n\tilde{\mu}_n^T\right] + \mathrm{tr}\left[W^T\Psi^{-1}W\left(\tilde{\mu}_n\tilde{\mu}_n^T + \tilde{\Sigma}_n\right)\right]\right)\right\}$$

Matrix Cookbook:

$$\frac{\partial}{\partial X}\mathrm{tr}\left(X^TBXC\right) = BXC + B^TXC^T \qquad \text{(eq. 117 p. 13)}$$

$$\frac{\partial}{\partial X}\mathrm{tr}\left(X^TA\right) = A \qquad \text{(eq. 103 p. 12)}.$$

$$\frac{\partial Q}{\partial W} = -\frac{1}{2}\sum_{n=1}^{N}\left(-2\Psi^{-1}x_n\tilde{\mu}_n^T + 2\Psi^{-1}W\left(\tilde{\mu}_n\tilde{\mu}_n^T + \tilde{\Sigma}_n\right)\right)$$

$$= +\Psi^{-1}\sum_{n=1}^{N}\left(x_n\tilde{\mu}_n^T\right) - \Psi^{-1}W\sum_{n=1}^{N}\left(\tilde{\mu}_n\tilde{\mu}_n^T + \tilde{\Sigma}_n\right)$$

$$= 0$$

$$\Leftrightarrow \boxed{W = \left(\sum_{n=1}^{N} x_n\tilde{\mu}_n^T\right)\left(\sum_{n=1}^{N}\tilde{\mu}_n\tilde{\mu}_n^T + \tilde{\Sigma}_n\right)^{-1}}$$

$$\frac{\partial Q}{\partial \psi} = \frac{\partial}{\partial \psi}\Bigg\{ -\frac{N}{2}\ln\det(\psi) - \frac{1}{2}\sum_{n=1}^{N}\Big[tr\left(\psi^{-1}x_n x_n^T\right) - 2tr\left(\psi^{-1}x_n \tilde{\mu}_n^T W^T\right) + tr\left(\psi^{-1}W\left[\tilde{\mu}_n\tilde{\mu}_n^T + \tilde{\Sigma}_n\right]W^T\right)\Big]\Bigg\}$$

Matrix Cookbook:

$$\frac{\partial \ln\det(X)}{\partial X} = \left(X^{-1}\right)^T$$

$$\frac{\partial tr(X^{-1}B)}{\partial X} = -\left(X^{-1}\right)^T B^T \left(X^{-1}\right)^T$$

$\hookrightarrow$ if X is diagonal, then $= -diag\left(\left(X^{-1}\right)^T B^T \left(X^{-1}\right)^T\right)$

where diag sets the off-diag elts to zero.

$$\frac{\partial Q}{\partial \psi} = -\frac{N}{2}\psi^{-1} - \frac{1}{2}\sum_{n=1}^{N}\left[-\psi^{-1}\left(x_n x_n^T - 2x_n\tilde{\mu}_n^T W^T + W\left[\tilde{\mu}_n\tilde{\mu}_n^T + \tilde{\Sigma}_n\right]W^T\right)\psi^{-1}\right]$$

$= 0$ (we multiply on the right & on the left by ψ)

$$\Leftrightarrow -\frac{N}{2}\psi + \frac{1}{2}\sum_{n=1}^{N} diag\left(x_n x_n^T - 2x_n\tilde{\mu}_n^T W^T + W\left[\tilde{\mu}_n\tilde{\mu}_n^T + \tilde{\Sigma}_n\right]W^T\right) = 0$$

$$\psi = \frac{1}{N}\sum_{n=1}^{N} diag\left(x_n x_n^T - 2x_n\tilde{\mu}_n^T W^T + W\left[\tilde{\mu}_n\tilde{\mu}_n^T + \tilde{\Sigma}_n\right]W^T\right)$$

$$\Psi = \frac{1}{N} \operatorname{diag} \left\{ \sum_n x_n x_n^T - 2\left(\sum_n x_n \tilde{\mu}_n^T\right) W^T + W\left(\sum_n \left[\tilde{\mu}_n \tilde{\mu}_n^T + \tilde{\Sigma}_n\right]\right) W^T \right\}$$

We can further inject the "new" W in front of the last term and we get:

$$\boxed{\Psi = \frac{1}{N} \operatorname{diag} \left\{ \sum_n x_n x_n^T - \left(\sum_n x_n \tilde{\mu}_n^T\right) W^T \right\}}$$