

1 Paired factor analysis (PFA) model

Let D_{nj} be the data corresponding to n -th sample and j -th feature, where n runs from 1 to N and j runs from 1 to J . Suppose these data come from a graph with K nodes (factors) and E edges. In the PFA set up $E = \frac{K(K-1)}{2}$.

Let us define latent variables Z and Λ . Z_n is a $(E + K) \times 1$ binary vector. We have a prior on $Z_{n,k_1,k_2,l}$, with the constraint that if $l = 0$, then k_1 and k_2 are both 0, and if k_1 and k_2 are non-zero, then l is 0.

$$\begin{aligned} Pr[Z_{n,k_1,k_2,0} = 1] &= \pi_{k_1,k_2} & k_1 < k_2 \\ Pr[Z_{n,0,0,l} = 1] &= \pi_l & l = 1, 2, \dots, K \end{aligned}$$

So, we assume that

$$\begin{aligned} \sum_{k_1 < k_2} \pi_{k_1 < k_2} + \sum_{l=1}^K \pi_l &= 1 \\ Pr(Z|\pi) &= \prod_{n=1}^N \prod_{k_1 < k_2}^E \pi_{k_1,k_2}^{z_{n,k_1,k_2,0}} \prod_{l=1}^K \pi_l^{z_{n,0,0,l}} \end{aligned}$$

Λ_n is a $Q \times 1$ binary vector, where Q is the cardinality of the set of values that q can take. For computational convenience we assume that q can take a finite set of values between 0 and 1, say $1/100, 2/100, \dots, 99/100, 1$.

We have a prior on Λ ,

$$\begin{aligned} Pr[\Lambda_{n,q} = 1] &= \delta_q \\ Pr(\Lambda|\delta) &= \prod_{n=1}^N \prod_{q=1}^Q \delta_q^{\lambda_{nq}} \end{aligned}$$

For data on the graph given edge (k_1, k_2) and position on edge q we assume the model

$$E[D_{nj} | Z_{n,k_1,k_2,0} = 1, \Lambda_{n,q} = 1, F] = qF_{k_1,j} + (1 - q)F_{k_2,j} \quad (1)$$

$$E[D_{nj} | Z_{0,0,l} = 1, F] = F_{l,j} \quad (2)$$

Where F is a $K \times J$ matrix of factors. Then we can marginalize over Z and Λ , assuming as a first pass that the data is Gaussian in its distribution,

$$Pr[D_n | Z_{n,k_1,k_2} = 1, \delta, F, s_{j=1,2,\dots,J}^2] = \sum_q \delta_q Pr[D_n | Z_{n,k_1,k_2} = 1, \Lambda_{n,q} = 1, F, s_{j=1,2,\dots,J}^2]$$

$$\begin{aligned} Pr [D_n | \pi, \delta, F, s_{j=1,2,\dots,J}^2] &= \sum_{k_1 < k_2} \pi_{k_1, k_2} Pr [D_n | Z_{n, k_1, k_2, 0} = 1, \delta, F, s_{j=1,2,\dots,J}^2] \\ &\quad + \sum_l \pi_l Pr [D_n | Z_{n, 0, 0, l} = 1, F, s_{j=1,2,\dots,J}^2] \end{aligned} \quad (3)$$

where s_j^2 is the residual variance of the j th feature.

We define the joint prior over the edges and the fraction of the edge represented as

$$\pi_{k_1, k_2, q} = \pi_{k_1, k_2} \delta_q \quad k_1 < k_2$$

The overall likelihood

$$L(\pi, F) = \prod_{n=1}^N Pr [D_n | \pi, F, s_{j=1,2,\dots,J}^2]$$

or we can write it as

$$\begin{aligned} L(\pi, F) &= \prod_{n=1}^N \left[\sum_{k_1 < k_2} \sum_{q=1}^Q \left[\pi_{k_1, k_2, q} \times \prod_{j=1}^J N(D_{nj}; qF_{k_1, g} + (1-q)F_{k_2, g}, s_j^2) \right] + \right. \\ &\quad \left. \sum_l \left[\pi_l \times \prod_{j=1}^J N(D_{nj}; F_{l, g}, s_j^2) \right] \right] \end{aligned} \quad (4)$$

And the log likelihood

$$\begin{aligned} \ln L(\pi, F) &= \sum_{n=1}^N \ln \left(\sum_{k_1 < k_2} \sum_q \left[\pi_{k_1, k_2, q} \times \prod_{j=1}^J N(D_{nj}; qF_{k_1, g} + (1-q)F_{k_2, g}, s_j^2) \right] + \right. \\ &\quad \left. \sum_{l=1}^K \pi_l \times \prod_{j=1}^J N(D_{nj}; F_{l, g}, s_j^2) \right) \end{aligned} \quad (5)$$

This is the quantity we want to maximize.

2 EM algorithm

2.1 E step

We assume that q can take a finite set of values between 0 and 1, say $1/100, 2/100, \dots, 90/100, 1$.

Suppose we have run upto m iterations. For the $(m+1)$ th iteration, we have

$$\begin{aligned} \delta_{n, k_1, k_2, 0, q}^{(m+1)} &= Pr [Z_{n, k_1, k_2, 0} = 1, \Lambda_{n, q} = 1 | \pi^{(m)}, F^{(m)}, s_{j=1,2,\dots,J}^{(m)}, D_n] \\ &\propto Pr [Z_{n, k_1, k_2, 0} = 1] Pr [\lambda_{n, q} = 1] Pr [D_n | \pi^{(m)}, F^{(m)}, s_{j=1,2,\dots,J}^{(m)}, Z_{n, k_1, k_2, 0} = 1, \lambda_{n, q} = 1] \\ &\propto \pi_{k_1, k_2, q}^{(m)} \prod_j N(D_{nj} | qF_{k_1, j}^{(m)} + (1-q)F_{k_2, j}^{(m)}, s_j^{(m)2}) \end{aligned}$$

$$\begin{aligned}
\delta_{n,0,0,l}^{(m+1)} &= Pr \left[Z_{n,0,0,l} = 1 | \pi^{(m)}, F^{(m)}, s_{j=1,2,\dots,J}^{(m)}, D_n \right] \\
&\propto Pr \left[Z_{n,0,0,l} = 1 \right] Pr \left[D_n | \pi^{(m)}, F^{(m)}, s_{j=1,2,\dots,J}^{(m)}, Z_{n,0,0,l} = 1 \right] \\
&\propto \pi_l^{(m)} \prod_j N \left(D_{nj} | F_{l,j}^{(m)}, s_j^{(m)2} \right)
\end{aligned}$$

where $s_j^{(m)2}$ is the residual variance of feature j .

We normalize δ so that

$$\sum_{k_1 < k_2} \sum_q \delta_{n,k_1,k_2,0,q}^{(m+1)} + \sum_{l=1}^K \delta_{n,0,0,l}^{(m+1)} = 1 \quad \forall n$$

We define

$$\begin{aligned}
\pi_{k_1,k_2,q}^{(m+1)} &= \frac{1}{N} \sum_{n=1}^N \delta_{n,k_1,k_2,0,q}^{(m+1)} \\
\pi_l^{(m+1)} &= \frac{1}{N} \sum_{n=1}^N \delta_{n,0,0,l}^{(m+1)}
\end{aligned}$$

We have therefore updated $\pi_{k_1,k_2,q}^{(m)}$ to $\pi_{k_1,k_2,q}^{(m+1)}$.

2.1.1 Variational EM - Model 1

In this set up, we assume prior distributions of π and δ as follows

$$Pr(\pi | \alpha_0) = C(\alpha_0) \prod_{k_1 < k_2} \pi_{k_1,k_2}^{\alpha_0-1} \prod_{l=1}^L \pi_l^{\alpha_0-1}$$

Similarly the prior distribution for δ is

$$Pr(\delta | \beta_0) = C(\beta_0) \prod_{q=1}^Q \delta_q^{\beta_0-1}$$

The likelihood above can be written as

$$\begin{aligned}
p(D | Z, \Lambda, F, s_{j=1,2,\dots,J}) &= \prod_{n=1}^N \left[\prod_{k_1 < k_2} \prod_{q=1}^Q \left[\prod_{j=1}^J N(D_{nj} | qF_{k_1,g} + (1-q)F_{k_2,g}, s_g^2) \right]^{\Lambda_{nq} Z_{n,k_1,k_2,0}} \right. \\
&\quad \left. \times \prod_l \left[\prod_{j=1}^J N(D_{nj} | F_{l,g}, s_g^2) \right]^{Z_{n,0,0,l}} \right] \quad (6)
\end{aligned}$$

The joint probability distribution is given by

$$p(D, Z, \Lambda, \pi, \delta | F, s_{j=1,2,\dots,J}, \alpha_0, \beta_0) = p(\pi | \alpha_0) p(\delta | \beta_0) p(\Lambda | \delta) p(Z | \pi) p(D | Z, \Lambda, F, s_{j=1,2,\dots,J})$$

We assume the following mean field variational distribution. In the first model, we assume the two latent variables Z and Λ are independent.

$$q(Z, \Lambda, \pi, \delta) = q(Z) q(\Lambda) q(\pi) q(\delta)$$

The variational distribution for Z

$$\begin{aligned} \ln q^*(Z) &= E_{\pi, \delta, \Lambda} [\ln p(\pi | \alpha_0) + \ln p(\delta | \beta_0) + \ln p(\Lambda | \delta) + \ln p(Z | \pi) + \ln p(D | Z, \Lambda, F, s_{j=1,2,\dots,J})] \\ &= E_{\pi, \delta, \Lambda} [\ln p(Z | \pi) + \ln p(D | Z, \Lambda, F, s_{j=1,2,\dots,J})] + \text{constant} \\ &= \sum_{n=1}^N \sum_{k_1 < k_2} z_{n,k_1,k_2} E_{\pi} [\ln(\pi_{k_1,k_2})] + \sum_{n=1}^N \sum_{k_1 < k_2} z_{n,k_1,k_2} \sum_q E_{\Lambda}(\lambda_{nq}) \left[- \sum_{j=1}^J \ln(s_j) \right. \\ &\quad \left. - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \end{aligned}$$

$$\begin{aligned} \ln q^*(\Lambda) &= E_{\pi, \delta, Z} [\ln p(\pi | \alpha_0) + \ln p(\delta | \beta_0) + \ln p(\Lambda | \delta) + \ln p(Z | \pi) + \ln p(D | Z, \Lambda, F, s_{j=1,2,\dots,J})] \\ &= E_{\pi, \delta, Z} [\ln p(\Lambda | \delta) + \ln p(D | Z, \Lambda, F, s_{j=1,2,\dots,J})] + \text{constant} \\ &= \sum_{n=1}^N \sum_{q=1}^Q \lambda_{n,q} E_{\delta} [\ln(\delta_q)] + \sum_{n=1}^N \sum_q \lambda_{nq} \sum_{k_1 < k_2} E_Z(z_{n,k_1,k_2}) \left[- \sum_{j=1}^J \ln(s_j) \right. \\ &\quad \left. - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \end{aligned}$$

So we get

$$q^*(Z) \propto \prod_{n=1}^N \prod_{k_1 < k_2} \rho_{n,k_1,k_2}^{Z_{n,k_1,k_2}}$$

where we define

$$\begin{aligned} \rho_{n,k_1,k_2} &\propto \exp \left(E_{\pi} [\ln(\pi_{k_1,k_2})] + \sum_q E_{\Lambda}(\lambda_{nq}) \left[- \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right) \\ \rho_{n,k_1,k_2} &\propto \exp \left(E_{\pi} [\ln(\pi_{k_1,k_2})] + \sum_q \nu_{nq} \left[- \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right) \end{aligned}$$

$$\rho_{n,k1,k2} \propto \exp \left(\psi_{a_{k1,k2}} - \psi \left(\sum_{k1 < k2} a_{k1,k2} \right) + \left[- \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

$\rho_{n,k1,k2}$ is normalized to sum to 1 for each n over $k1$ and $k2$.

We also get

$$q^*(\Lambda) \propto \prod_{n=1}^N \prod_{q=1}^Q \nu_{nq}^{\Lambda_{nq}}$$

where

$$\nu_{nq} \propto \exp \left(E_\delta [\ln(\delta_q)] + \sum_{k1 < k2} E_Z(z_{n,k1,k2}) \left[- \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

$$\nu_{nq} \propto \exp \left(E_\delta [\ln(\delta_q)] + \sum_{k1 < k2} \rho_{n,k1,k2} \left[- \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

$$\nu_{nq} \propto \exp \left(\psi(b_q) - \psi \left(\sum_{q=1}^Q b_q \right) + \left[- \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2} \right] \right)$$

ν_{nq} are normalized to sum to 1.

We can also derive variational distributions similarly for π and δ .

$$\begin{aligned} \ln q^*(\pi) &= E_{\Lambda,Z,\delta} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\ &= E_Z [\ln p(Z|\pi)] + \ln p(\pi|\alpha_0) + \text{constant} \\ &= \sum_{n=1}^N \sum_{k1 < k2} E(z_{n,k1,k2}) \ln \pi_{k1,k2} + (\alpha_0 - 1) \sum_{k1 < k2} \ln \pi_{k1,k2} \\ &= \sum_{k1 < k2} \left[\sum_{n=1}^N \rho_{n,k1,k2} + (\alpha_0 - 1) \right] \ln \pi_k \end{aligned}$$

We define

$$a_{k1,k2} = \alpha_0 + \sum_{n=1}^N \rho_{n,k1,k2}$$

$$q^*(\pi) = \text{Dir}(\pi|a)$$

$$\begin{aligned}
\ln q^*(\delta) &= E_{\Lambda, Z, \pi} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\
&= E_{\Lambda} [\ln p(\Lambda|\delta)] + \ln p(\delta|\beta_0) + \text{constant} \\
&= \sum_{n=1}^N \sum_{q=1}^Q E(\lambda_{n,q}) \ln \delta_q + (\beta_0 - 1) \sum_{q=1}^Q \ln \delta_q \\
&= \sum_{q=1}^Q \left[\sum_{n=1}^N \nu_{n,q} + (\beta_0 - 1) \right] \ln \delta_q
\end{aligned}$$

We define

$$b_q = \beta_0 + \sum_{n=1}^N \nu_{n,q}$$

$$q^*(\delta) = \text{Dir}(\delta|b)$$

We alternate between the Variational E and M steps, E steps being the ones where we compute the responsibilities $\rho_{n,k1,k2}$ and $\nu_{n,q}$ and the M step is where we update the variational distribution of the parameters π and δ .

We can start with $a = \alpha_0$ and $b = \beta_0$. We can then estimate $\rho_{n,k1,k2}$ and also ν_{nq} and then the product of these two terms to get new responsibility

$$\delta_{n,k1,k2,q} = \rho_{n,k1,k2} \nu_{nq}$$

and we use this $\delta_{n,k1,k2,q}$ as the responsibility for the M-step of the original EM updates.

2.1.2 Variational EM - Model 2

In model 2, we do not assume independence of the latent variables Z and Λ and instead estimate their joint variational distribution.

$$q(Z, \Lambda, \pi, \delta) = q(Z, \Lambda)q(\pi)q(\delta)$$

$$\begin{aligned}
\ln q^*(Z, \Lambda) &= E_{\pi, \delta} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \quad (7) \\
&= E_{\pi, \delta, \Lambda} [\ln p(Z|\pi) + \ln p(\Lambda|\delta) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] + \text{constant} \\
&= \sum_{n=1}^N \sum_{k_1 < k_2} \sum_{q=1}^Q z_{n,k_1,k_2,0} \lambda_{nq} E_{\pi} [\ln(\pi_{k_1,k_2})] + \sum_{n=1}^N \sum_{q=1}^Q \sum_{k_1 < k_2} \lambda_{n,q} z_{n,k_1,k_2,0} E_{\delta} [\ln(\delta_q)] \\
&\quad + \sum_{n=1}^N \sum_{k_1 < k_2} \sum_q z_{n,k_1,k_2,0} \lambda_{nq} \left[-\sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \\
&\quad + \sum_{n=1}^N \sum_l z_{n,0,0,l} \left[-\sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - F_{l,j})^2}{2s_j^2} \right] \\
&\quad + \sum_{n=1}^N \sum_{l=1}^K z_{n,0,0,l} E_{\pi} [\ln(\pi_l)]
\end{aligned}$$

From here one can get

$$q^*(Z, \Lambda) \propto \prod_{n=1}^N \left[\prod_{k_1 < k_2} \prod_{q=1}^Q \delta_{n,k_1,k_2,0,q}^{Z_{n,k_1,k_2,0,q}} \prod_{l=1}^K \delta_{n,0,0,l,q}^{Z_{n,0,0,l,q}} \right]$$

then

$$\begin{aligned}
\delta_{n,k_1,k_2,0,q} &\propto \exp \left(E_{\pi} [\ln(\pi_{k_1,k_2})] + E_{\delta} [\ln(\delta_q)] \right. \\
&\quad \left. + \left[-\sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right) \quad (8)
\end{aligned}$$

$$\begin{aligned}
\delta_{n,k_1,k_2,0,q} &\propto \exp \left(\psi(a_{k_1,k_2,0}) - \psi\left(\sum_l a_{0,0,l} + \sum_{k_1 < k_2} a_{k_1,k_2,0}\right) + \psi(b_q) - \psi\left(\sum_{q=1}^Q b_q\right) \right. \\
&\quad \left. \left[-\sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right) \quad (9)
\end{aligned}$$

$$\delta_{n,0,0,l} \propto \exp \left(E_{\pi} [\ln(\pi_l)] + \left[-\sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - F_{l,j})^2}{2s_j^2} \right] \right) \quad (10)$$

$$\delta_{n,0,0,l} \propto \exp \left(\psi_{a_{0,0,l}} - \psi\left(\sum_l a_{0,0,l} + \sum_{k_1 < k_2} a_{k_1,k_2,0}\right) + \left[-\sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - F_{l,j})^2}{2s_j^2} \right] \right) \quad (11)$$

We can also derive variational distributions similarly for π and δ .

$$\begin{aligned}
\ln q^*(\pi) &= E_{\Lambda, Z, \delta} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\
&= E_Z [\ln p(Z|\pi)] + \ln p(\pi|\alpha_0) + \text{constant} \\
&= \sum_{n=1}^N \sum_{k1 < k2} E(z_{n,k1,k2,0}) \ln \pi_{k1,k2} + \sum_{n=1}^N \sum_{l=1}^K E(z_{n,0,0,l}) \ln \pi_l + (\alpha_0 - 1) \sum_{k1 < k2} \ln \pi_{k1,k2} \\
&= \sum_{k1 < k2} \left[\sum_{n=1}^N \rho_{n,k1,k2,0} + (\alpha_0 - 1) \right] \ln \pi_{k1,k2} + \sum_{l=1}^K \left[\sum_{n=1}^N \rho_{n,0,0,l} + (\alpha_0 - 1) \right] \ln \pi_l
\end{aligned}$$

We define

$$\begin{aligned}
a_{k1,k2,0} &= \alpha_0 + \sum_{n=1}^N \rho_{n,k1,k2,0} \\
a_{0,0,l} &= \alpha_0 + \sum_{n=1}^N \rho_{n,0,0,l} \\
q^*(\pi) &= \text{Dir}(\pi|a)
\end{aligned}$$

$$\begin{aligned}
\ln q^*(\delta) &= E_{\Lambda, Z, \pi} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\
&= E_{\Lambda} [\ln p(\Lambda|\delta)] + \ln p(\delta|\beta_0) + \text{constant} \\
&= \sum_{n=1}^N \sum_{q=1}^Q E(\lambda_{n,q}) \ln \delta_q + (\beta_0 - 1) \sum_{q=1}^Q \ln \delta_q \\
&= \sum_{q=1}^Q \left[\sum_{n=1}^N \nu_{n,q} + (\beta_0 - 1) \right] \ln \delta_q
\end{aligned}$$

We define

$$\begin{aligned}
b_q &= \beta_0 + \sum_{n=1}^N \nu_{n,q} \\
q^*(\delta) &= \text{Dir}(\delta|b)
\end{aligned}$$

The updates for π and δ are same as before. Here also we have the same way of initializing π and δ first, then use $a_{k1,k2} = \alpha_0$ and $b_q = \beta_0$ to begin with and estimate $\delta_{n,k1,k2,q}$. Then use the $\delta_{n,k1,k2,q}$ to update $a_{k1,k2}$ and b_q and proceed in this way. In this case, we do not assume independence of the Λ and Z variational distributions, so this model is more generalized.

2.2 M step

We take the expectation of this quantity with respect to $[Z, \lambda|D, \theta^{(m)}]$.

$$Q(\theta|\theta^{(m)}) \propto - \sum_{n=1}^N \sum_{k_1 < k_2} \sum_{q=1}^Q \delta_{n,k_1,k_2,0,q}^{(m+1)} \sum_j \left[\log s_j^{(m+1)} + \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^{(m+1)^2}} \right] \quad (12)$$

$$- \sum_{n=1}^N \sum_{l=1}^K \sum_{q=1}^Q \delta_{n,0,0,l,q}^{(m+1)} \sum_j \left[\log s_j^{(m+1)} + \frac{(D_{nj} - F_{l,j})^2}{2s_j^{(m+1)^2}} \right] \quad (13)$$

We try to maximize this quantity with respect to F , So, we can take derivative with respect to F and try to solve the resulting normal equation.

This equation, conditional on $[Z, \lambda|D, \theta^{(m)}]$, can be written as

$$D_{N \times J} = L_{N \times K} F_{K \times J} + E_{N \times J} \quad (14)$$

where

$$e_{nj} \sim N(0, s_j^2)$$

We define

$$D'_{nj} := \frac{D_{nj}}{s_j}$$

If we consider finding the factors on a feature-by-feature basis, we do not need to worry about s_j .

$$L_{nk} = \begin{cases} q \text{ or } (1-q) & \lambda_n = q \text{ } Z_{n,k,*,0} = 1 \text{ or } Z_{n,*,k,0} = 1 \\ 1 & Z_{n,0,0,k} = 1 \\ 0 & \text{o.w.} \end{cases}$$

We have

$$E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}] = \sum_q \sum_{k_2 > k} q \delta_{n,k,k_2,0,q}^{(m+1)} + \sum_q \sum_{k_1 < k} (1-q) \delta_{n,k_1,k,0,q}^{(m+1)} + \delta_{n,0,0,k}^{(m+1)}$$

$$E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}^2] = \sum_q \sum_{k_2 > k} q^2 \delta_{n,k,k_2,q}^{(m+1)} + \sum_q \sum_{k_1 < k} (1-q)^2 \delta_{n,k_1,k,q}^{(m+1)} + \delta_{n,0,0,k}^{(m+1)}$$

Also for any $k \neq l$,

$$E_{Z,\lambda|D,\theta^{(m)}} [L_{nk} L_{nl}] = \sum_q q(1-q) \delta_{n,k,l,q}^{(m+1)}$$

We use these to solve for the equation

$$\left[E_{Z,\lambda|D,\theta^{(m)}} (L^T L) \right] F \approx \left[E_{Z,\lambda|D,\theta^{(m)}} (L) \right]^T D$$

The solution therefore is

$$F \approx \left[E_{Z,\lambda|D,\theta^{(m)}} (L^T L) \right]^{-1} \left[E_{Z,\lambda|D,\theta^{(m)}} (L) \right]^T D$$

For $W = L^T L$

$$W_{kl} = \sum_n L_{kn} L_{nl}$$

$$E_{Z,\lambda|D,\theta^{(m)}} (W_{kl}) = \sum_n E_{Z,\lambda|D,\theta^{(m)}} (L_{nk} L_{nl})$$

We use the definition of $E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}^2]$ and $E_{Z,\lambda|D,\theta^{(m)}} [L_{nk} L_{nl}]$ from above to solve F .

In the same way as we computed F by solving for the normal equation obtained from taking derivative of the function $Q(\theta|\theta^{(m)})$, we take derivative of the latter with respect to s_j^2 to obtain EM updates of the residual variance terms. Taking the derivative, we obtain the estimate as

$$\begin{aligned} \widehat{s_j^{(m+1)}}^2 &= \frac{1}{N} \sum_{n=1}^N \sum_{k_1 < k_2} \sum_q \delta_{n,k_1,k_2,0,q}^{(m+1)} (D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2 \\ &\quad + \frac{1}{N} \sum_{n=1}^N \sum_{l=1}^K \delta_{n,0,0,l,q}^{(m+1)} (D_{nj} - F_{l,j})^2 \quad (15) \end{aligned}$$

where the F are the estimated values of the factors from the previous step.

We then continue this procedure described above for multiple iterations.