

Algorithm for “Principal component models for sparse functional data”

1 Appendix

In this section we provide details of the Reduced Rank fitting algorithm. Steps 1 and 2 make up the M-step and Step 3 makes up the E-step.

1. Given current estimates for α_i , θ_0 and θ we estimate σ^2 and D as

$$\begin{aligned}
 \hat{\sigma}^2 &= \frac{1}{\sum n_i} \sum_{i=1}^N E[\epsilon_i^T \epsilon_i | \mathbf{Y}_i] \\
 &= \frac{1}{\sum n_i} \sum_{i=1}^N E[(\mathbf{Y}_i - B_i \hat{\theta}_0 - B_i \hat{\theta} \alpha_i)^T (\mathbf{Y}_i - B_i \hat{\theta}_0 - B_i \hat{\theta} \alpha_i) | \mathbf{Y}_i] \\
 &= \frac{1}{\sum n_i} \sum_{i=1}^N \left((\mathbf{Y}_i - B_i \hat{\theta}_0 - B_i \hat{\theta} \hat{\alpha}_i)^T (\mathbf{Y}_i - B_i \hat{\theta}_0 - B_i \hat{\theta} \hat{\alpha}_i) \right. \\
 &\quad \left. + \text{trace}[B_i \hat{\theta} (\hat{D}^{-1} + \hat{\theta}^T B_i^T B_i \hat{\theta} / \hat{\sigma}^2)^{-1} \hat{\theta}^T B_i^T] \right) \tag{1}
 \end{aligned}$$

$$\hat{D}_{jj} = \frac{1}{N} \sum_{i=1}^N E[\alpha_{ij}^2 | \mathbf{Y}_i] = \frac{1}{N} \sum_{i=1}^N \left(\hat{\alpha}_{ij}^2 + (\hat{D}^{-1} + \hat{\theta}^T B_i^T B_i \hat{\theta} / \hat{\sigma}^2)^{-1}_{jj} \right) \tag{2}$$

Equations (1) and (2) derive from the facts that

$$E(X^2|Y) = (E(X|Y))^2 + \text{Var}(X|Y) \tag{3}$$

and

$$\alpha_i | \mathbf{Y}_i \sim N \left((\sigma^2 D^{-1} + \theta^T B_i^T B_i \theta)^{-1} \theta B_i^T (\mathbf{Y}_i - B_i \theta_0), (D^{-1} + \theta^T B_i^T B_i \theta / \sigma^2)^{-1} \right) \tag{4}$$

2. Given current estimates for σ^2 , D and α_i we estimate θ and θ_0 by minimizing

$$\sum_{i=1}^N \left[(\mathbf{Y}_i - B_i \hat{\theta}_0 - B_i \hat{\theta} \hat{\alpha}_i)^T (\mathbf{Y}_i - B_i \hat{\theta}_0 - B_i \hat{\theta} \hat{\alpha}_i) \right] \tag{5}$$

Minimizing (5) involves a second iterative procedure, in which each column of θ is estimated

separately holding all other columns fixed. First notice that

$$\begin{aligned} & \sum_{i=1}^N \|\mathbf{Y}_i - B_i \boldsymbol{\theta}_0 - B_i \boldsymbol{\theta} \boldsymbol{\alpha}_i\|^2 \\ &= \sum_{i=1}^N \|(\mathbf{Y}_i - B_i \boldsymbol{\theta} \boldsymbol{\alpha}_i) - B_i \boldsymbol{\theta}_0\|^2 \end{aligned}$$

so the estimate for $\boldsymbol{\theta}_0$ is

$$\hat{\boldsymbol{\theta}}_0 = \left(\sum_{i=1}^N B_i^T B_i \right)^{-1} \sum_{i=1}^N B_i^T (\mathbf{Y}_i - B_i \hat{\boldsymbol{\theta}} \hat{\boldsymbol{\alpha}}_i) \quad (6)$$

To estimate the columns of $\boldsymbol{\theta}$ we note that

$$\begin{aligned} & \sum_{i=1}^N \|\mathbf{Y}_i - B_i \boldsymbol{\theta}_0 - B_i \boldsymbol{\theta} \boldsymbol{\alpha}_i\|^2 \\ &= \sum_{i=1}^N \|(\mathbf{Y}_i - B_i \boldsymbol{\theta}_0 - \alpha_{i2} B_i \boldsymbol{\theta}_2 - \cdots - \alpha_{ik} B_i \boldsymbol{\theta}_k) - \alpha_{i1} B_i \boldsymbol{\theta}_1\|^2 \end{aligned}$$

Therefore the estimate for $\boldsymbol{\theta}_1$ is

$$\hat{\boldsymbol{\theta}}_1 = \left(\sum_{i=1}^N \hat{\alpha}_{i1}^2 B_i^T B_i \right)^{-1} \sum_{i=1}^N B_i^T (\hat{\alpha}_{i1} (\mathbf{Y}_i - B_i \hat{\boldsymbol{\theta}}_0) - \hat{\alpha}_{i1} \hat{\alpha}_{i2} B_i \hat{\boldsymbol{\theta}}_2 - \cdots - \hat{\alpha}_{i1} \hat{\alpha}_{ik} B_i \hat{\boldsymbol{\theta}}_k) \quad (7)$$

We repeat this procedure for each column of $\boldsymbol{\theta}$ and iterate until there is no further change.

3. The E-step consists of predicting $\boldsymbol{\alpha}_i$ and $\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^T$.

$$\hat{\boldsymbol{\alpha}}_i = E(\boldsymbol{\alpha}_i | \mathbf{Y}_i, \hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2, \hat{D}) = (\hat{\sigma}^2 \hat{D}^{-1} + \hat{\boldsymbol{\theta}}^T B_i^T B_i \hat{\boldsymbol{\theta}})^{-1} \hat{\boldsymbol{\theta}}^T B_i^T (\mathbf{Y}_i - B_i \hat{\boldsymbol{\theta}}_0) \quad (8)$$

$$\widehat{\boldsymbol{\alpha}}_i \widehat{\boldsymbol{\alpha}}_i^T = E(\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^T | \mathbf{Y}_i, \hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\theta}}, \hat{\sigma}^2, \hat{D}) = \hat{\boldsymbol{\alpha}}_i \hat{\boldsymbol{\alpha}}_i^T + (\hat{D}^{-1} + \hat{\boldsymbol{\theta}}^T B_i^T B_i \hat{\boldsymbol{\theta}} / \hat{\sigma}^2)^{-1} \quad (9)$$

Both predictions make use of equations (3) and (4).

4. We then return to Step 1 and repeat until we reach convergence.
5. The matrix $\boldsymbol{\theta}$ produced by this procedure will not be orthogonal. We orthogonalize it by producing the reduced rank estimate for Γ ,

$$\hat{\Gamma} = \hat{\boldsymbol{\theta}} \hat{D} \hat{\boldsymbol{\theta}}^T \quad (10)$$

and setting $\boldsymbol{\theta}$ equal to the first k eigenvectors of $\hat{\Gamma}$.