Supplementary Material for Bayesian Functional Principal Components Analysis via Variational Message Passing

A Proof of Theorem 2.1

We first note that

$$y_i(t) - \mu(t) = \sum_{l=1}^{L} z_{il} h_l(t), \quad i = 1, \ldots, n. \tag{A.1}$$

The existence of an orthonormal eigenbasis $\psi_1, \ldots, \psi_L$ can be established via Gram-Schmidt orthogonalization. We first set

$$\phi_1 \equiv h_1, \quad \phi_j \equiv h_j - \sum_{l=1}^{j-1} \frac{\langle \phi_l, h_j \rangle}{||\phi_l||^2} \phi_l, \quad j = 2, \ldots, L.$$

Next, set

$$\phi_j^* = \frac{\phi_j}{||\phi_j||}, \quad j = 1, \ldots, L.$$

Then $\phi_1^*, \ldots, \phi_L^*$ form an orthonormal basis for the span of $h_1, \ldots, h_L$. Therefore, (A.1) can be re-written as

$$y_i(t) - \mu(t) = \sum_{l=1}^{L} \iota_{il} \phi_l^*(t), \quad i = 1, \ldots, n,$$

where

$$\iota_{il} \equiv z_{il}||\phi_l|| + \sum_{j=l+1}^{L} z_{ij}\frac{\langle \phi_l, h_j\rangle}{||\phi_l||}, \quad l = 1, \ldots, L-1, \quad \iota_{iL} \equiv z_{iL}||\phi_L||.$$

Note that $\iota_{i1}, \ldots, \iota_{iL}$ are correlated.

Now, define $\boldsymbol{\iota}_i \equiv (\iota_{i1}, \ldots, \iota_{iL})^\intercal$, $i = 1, \ldots, n$. Since the curves $y_1, \ldots, y_n$ are random observations of a Gaussian process, we have

$$\boldsymbol{\iota}_i \overset{\text{ind.}}{\sim} \mathrm{N}(0, \boldsymbol{\Sigma}_\iota), \quad i = 1, \ldots, n.$$

Next, establish the eigendecomposition of Σ_ι, such that $\boldsymbol{\Sigma}_\iota = \boldsymbol{Q}_\iota \boldsymbol{\Lambda}_\iota \boldsymbol{Q}_\iota^\intercal$, where $\boldsymbol{\Lambda}_\iota$ is a diagonal matrix consisting of the eigenvalues of $\boldsymbol{\Sigma}_\iota$ in descending order, and the columns of $\boldsymbol{Q}_\iota$ are the corresponding eigenvectors. Then, it can be easily seen that

$$\boldsymbol{\zeta}_i \equiv \boldsymbol{Q}_\iota^\intercal \boldsymbol{\iota}_i \overset{\text{ind.}}{\sim} \mathrm{N}(0, \boldsymbol{\Lambda}_\iota), \quad i = 1, \ldots, n.$$

That is, the elements of $\boldsymbol{\zeta}_i$ are uncorrelated and $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n$ are independent.

Next, define the eigenvectors of $\boldsymbol{\Sigma}_\iota$ as $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_L$, such that $\boldsymbol{Q} = [\ \boldsymbol{q}_1 \quad \cdots \quad \boldsymbol{q}_L\ ]$. Furthermore, define the elements of each of the eigenvectors such that $\boldsymbol{q}_l = (q_{l1}, \ldots, q_{lL})^\intercal$, $l = 1, \ldots, L$. Then, set

$$\psi_l \equiv \sum_{j=1}^{L} q_{jl}\phi_j^*, \quad l = 1, \ldots, L.$$

The orthonormality of $\psi_1, \ldots, \psi_L$ is easily verified:

$$\begin{aligned}
\langle \psi_l, \psi_j \rangle &= \left\langle \sum_{m=1}^{L} q_{ml}\phi_m^*, \sum_{k=1}^{L} q_{kj}\phi_k^* \right\rangle = \sum_{m=1}^{L}\sum_{k=1}^{L} q_{ml}q_{kj}\langle \phi_m^*, \phi_k^*\rangle \\
&= \sum_{m=1}^{L} q_{ml}q_{mj} = \boldsymbol{q}_l^\intercal \boldsymbol{q}_j = \mathbb{I}(l = j),
\end{aligned}$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Finally, we have

$$
\begin{aligned}
y_i(t) - \mu(t) &= \sum_{l=1}^{L} \iota_{il}\phi_l^*(t) = \sum_{l=1}^{L}\sum_{j=1}^{L} \zeta_{ij} q_{lj}\phi_l^*(t) \\
&= \sum_{j=1}^{L} \zeta_{ij} \sum_{l=1}^{L} q_{lj}\phi_l^*(t) = \sum_{j=1}^{L} \zeta_{ij}\ \psi_j(t).
\end{aligned}
$$

This decomposition is unique up to a change of sign.

B Exponential Family Form

We first describe the exponential family form for the normal distribution. For a $d \times 1$ multivariate normal random vector $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the probability density function of $\boldsymbol{x}$ can be shown to satisfy

$$
p(\boldsymbol{x}) = \exp\left\{ \boldsymbol{T}_{\text{vec}}(\boldsymbol{x})^\intercal \boldsymbol{\eta}_{\text{vec}} - A_{\text{vec}}(\boldsymbol{\eta}_{\text{vec}}) - \frac{d}{2}\log(2\pi) \right\}, \tag{B.1}
$$

where $\boldsymbol{T}_{\text{vec}}(\boldsymbol{x}) \equiv \{\boldsymbol{x}^\intercal, \text{vec}(\boldsymbol{x}\boldsymbol{x}^\intercal)^\intercal\}^\intercal$ is the vector of sufficient statistics and $\boldsymbol{\eta}_{\text{vec}} \equiv (\boldsymbol{\eta}_{\text{vec},1}^\intercal, \boldsymbol{\eta}_{\text{vec},2}^\intercal)^\intercal \equiv [(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^\intercal, -\frac{1}{2}\{\text{vec}(\boldsymbol{\Sigma}^{-1})\}^\intercal]^\intercal$ is the natural parameter vector. The function $A_{\text{vec}}(\boldsymbol{\eta}_{\text{vec}}) = -\frac{1}{4}\boldsymbol{\eta}_{\text{vec},1}^\intercal\{\text{vec}^{-1}(\boldsymbol{\eta}_{\text{vec},2})\}^{-1}\boldsymbol{\eta}_{\text{vec},1} - \frac{1}{2}\log|-2\,\text{vec}^{-1}(\boldsymbol{\eta}_{\text{vec},2})|$ is the log-partition function. The inverse mapping of the natural parameter vector is (Wand, 2017, equation S.4)

$$
\boldsymbol{\mu} = -\frac{1}{2}\left\{\text{vec}^{-1}(\boldsymbol{\eta}_{\text{vec},2})\right\}^{-1}\boldsymbol{\eta}_{\text{vec},1} \quad \text{and} \quad \boldsymbol{\Sigma} = -\frac{1}{2}\left\{\text{vec}^{-1}(\boldsymbol{\eta}_{\text{vec},2})\right\}^{-1}. \tag{B.2}
$$

We will refer to the representation of the multivariate normal probability density function in (B.1) as the *vec-based representation*.

Alternatively, a more storage-economical representation of the multivariate normal probability density function is the *vech-based representation*:

$$
p(\boldsymbol{x}) = \exp\left\{ \boldsymbol{T}_{\text{vech}}(\boldsymbol{x})^\intercal \boldsymbol{\eta}_{\text{vech}} - A_{\text{vech}}(\boldsymbol{\eta}_{\text{vech}}) - \frac{d}{2}\log(2\pi) \right\},
$$

where the vector of sufficient statistics, the natural parameter vector and the log-partition function are, $\boldsymbol{T}_{\text{vech}}(\boldsymbol{x}) \equiv \{\boldsymbol{x}^\intercal, \text{vec}(\boldsymbol{x}\boldsymbol{x}^\intercal)^\intercal\}^\intercal$, $\boldsymbol{\eta}_{\text{vech}} \equiv (\boldsymbol{\eta}_{\text{vech},1}^\intercal, \boldsymbol{\eta}_{\text{vech},2}^\intercal)^\intercal \equiv [(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^\intercal, -\frac{1}{2}\boldsymbol{D}_d^\intercal\{\text{vec}(\boldsymbol{\Sigma}^{-1})\}^\intercal]^\intercal$ and $A_{\text{vech}}(\boldsymbol{\eta}_{\text{vech}}) = -\frac{1}{4}\boldsymbol{\eta}_{\text{vech},1}^\intercal\{\text{vec}^{-1}(\boldsymbol{D}_d^{+\intercal}\boldsymbol{\eta}_{\text{vech},2})\}^{-1}\boldsymbol{\eta}_{\text{vech},1} -$

$\frac{1}{2}\log|-2\,\mathrm{vec}^{-1}(\boldsymbol{D}_d^{+\intercal}\boldsymbol{\eta}_{\mathrm{vech},2})|$, respectively. The inverse mapping of the natural parameter vector under the vech-based representation is

$$\boldsymbol{\mu} = -\frac{1}{2}\left\{\mathrm{vec}^{-1}(\boldsymbol{D}_d^{+\intercal}\boldsymbol{\eta}_{\mathrm{vech},2})\right\}^{-1}\boldsymbol{\eta}_{\mathrm{vech},1} \quad \text{and} \quad \boldsymbol{\Sigma} = -\frac{1}{2}\left\{\mathrm{vec}^{-1}(\boldsymbol{D}_d^{+\intercal}\boldsymbol{\eta}_{\mathrm{vech},2})\right\}^{-1}. \tag{B.3}$$

The other major distribution within the exponential family that is pivotal for this article is the inverse-χ^2 distribution. A random variable x has an inverse-χ^2 distribution with shape parameter $\xi > 0$ and scale parameter $\lambda > 0$ if the probability density function of x is

$$p(x) = \frac{(\lambda/2)^{\xi/2}}{\Gamma(\xi/2)}x^{-(\xi+2)/2}\exp\left(-\frac{\lambda}{2x}\right)\mathbb{I}(x>0),$$

where the vector of sufficient statistics, the natural parameter vector and the log-partition function are $\boldsymbol{T}(x) \equiv (\log(x), 1/x)^\intercal$, $\boldsymbol{\eta} = (\eta_1, \eta_2)^\intercal = \{-\frac{1}{2}(\xi+2), -\frac{\lambda}{2}\}^\intercal$ and $A(\boldsymbol{\eta}) \equiv \log\{\Gamma(\xi/2)\} - \frac{\xi}{2}\log(\lambda/2)$, respectively. Note that $\Gamma(z) \equiv \int_0^\infty u^{z-1}e^u du$ is the gamma function, $\mathbb{I}(\cdot)$ is the indicator function, $\zeta > 0$ is the scale parameter and $\lambda > 0$ is the shape parameter. The inverse mapping of the natural parameter vector is $\xi = -2\eta_1 - 2$ and $\lambda = -2\eta_2$.

C Proof of Lemma 5.1

To prove (5.4), we first note that posterior curve estimates from the VMP algorithm satisfy

$$\begin{aligned}
\widehat{y}_i(\boldsymbol{t}_g) &= \boldsymbol{C}_g\,\mathbb{E}_q(\boldsymbol{\nu}_\mu) + \sum_{l=1}^{L}\mathbb{E}_q(\zeta_{il})\boldsymbol{C}_g\,\mathbb{E}_q(\boldsymbol{\nu}_{\psi_l}) \\
&= \widehat{\mu}(\boldsymbol{t}_g) + \sum_{l=1}^{L}\mathbb{E}_q(\zeta_{il})\,\mathbb{E}_q\{\psi_l(\boldsymbol{t}_g)\} \\
&= \widehat{\mu}(\boldsymbol{t}_g) + \boldsymbol{\Psi}\,\mathbb{E}_q(\boldsymbol{\zeta}_i) \\
&= \widehat{\mu}(\boldsymbol{t}_g) + \boldsymbol{U}_\psi\boldsymbol{D}_\psi\boldsymbol{V}_\psi^\intercal\,\mathbb{E}_q(\boldsymbol{\zeta}_i) \\
&= \widehat{\mu}(\boldsymbol{t}_g) + \boldsymbol{U}_\psi\boldsymbol{Q}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^\intercal\boldsymbol{D}_\psi\boldsymbol{V}_\psi^\intercal\,\mathbb{E}_q(\boldsymbol{\zeta}_i) \\
&= \widehat{\mu}(\boldsymbol{t}_g) + \ddot{\boldsymbol{\Psi}}\ddot{\boldsymbol{\zeta}}_i,
\end{aligned} \tag{C.1}$$

where $\overset{\bullet}{\boldsymbol{\zeta}}_i \equiv (\dot{\zeta}_{11}, \dots, \dot{\zeta}_{1L})^\intercal$, $i = 1, \dots, n$. Next, define

$$\boldsymbol{Y} \equiv \left[\widehat{y}_1(\boldsymbol{t}_g) \quad \dots \quad \widehat{y}_n(\boldsymbol{t}_g)\right]$$

Then, (C.1) implies

$$\boldsymbol{Y} - \widehat{\mu}(\boldsymbol{t}_g)\boldsymbol{1}_N^\intercal = \overset{\bullet}{\boldsymbol{\Psi}}\overset{\bullet}{\boldsymbol{\Xi}}^\intercal.$$

Now, let $\boldsymbol{c}$ be the $L \times 1$ vector, with $||\dot{\psi}_l||$ as the lth entry, $l = 1, \dots, L$. Furthermore, let $1/\boldsymbol{c}$ be the $L \times 1$ vector, with $1/||\dot{\psi}_l||$ as the lth entry, $l = 1, \dots, L$. Recall that we can approximate these values through numerical integration. Then,

$$\boldsymbol{Y} - \widehat{\mu}(\boldsymbol{t}_g)\boldsymbol{1}_N^\intercal = \overset{\bullet}{\boldsymbol{\Psi}}\,\mathrm{diag}(1/\boldsymbol{c})\,\mathrm{diag}(\boldsymbol{c})\overset{\bullet}{\boldsymbol{\Xi}}^\intercal.$$

It is easy to see that this implies (5.4).

D Proof of Proposition 5.1

The independence of $\widehat{\boldsymbol{\zeta}}_1, \dots, \widehat{\boldsymbol{\zeta}}_n$ is a consequence of the independence assumption in (3.1). Let $\boldsymbol{c}$ and $1/\boldsymbol{c}$ retain their definitions from Appendix C. Then, note that

$$\widehat{\boldsymbol{\zeta}}_i = \mathrm{diag}(\boldsymbol{c})\overset{\bullet}{\boldsymbol{\zeta}}_i = \mathrm{diag}(\boldsymbol{c})\boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^\intercal\boldsymbol{D}_\psi\boldsymbol{V}_\psi^\intercal\,\mathbb{E}_q(\boldsymbol{\zeta}_i).$$

Let $\boldsymbol{m}_{\widehat{\zeta}}$ be the sample mean vector of $\widehat{\boldsymbol{\zeta}}_1, \dots, \widehat{\boldsymbol{\zeta}}_n$. Then,

$$\begin{aligned}
\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{m}_{\widehat{\zeta}} &= \mathrm{diag}(\boldsymbol{c})\boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^\intercal\boldsymbol{D}\boldsymbol{V}^\intercal\,\mathbb{E}_q(\boldsymbol{\zeta}_i) - \boldsymbol{m}_{\widehat{\zeta}} \\
&= \mathrm{diag}(\boldsymbol{c})\boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^\intercal\{\boldsymbol{D}\boldsymbol{V}^\intercal\,\mathbb{E}_q(\boldsymbol{\zeta}_i) - \boldsymbol{Q}\boldsymbol{\Lambda}^{1/2}\,\mathrm{diag}(1/\boldsymbol{c})\boldsymbol{m}_{\widehat{\zeta}}\} \\
&= \mathrm{diag}(\boldsymbol{c})\boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^\intercal\{\boldsymbol{D}\boldsymbol{V}^\intercal\,\mathbb{E}_q(\boldsymbol{\zeta}_i) - \boldsymbol{m}_\zeta\}
\end{aligned}$$

where $\boldsymbol{m}_\zeta \equiv \boldsymbol{Q}\boldsymbol{\Lambda}^{1/2}\,\mathrm{diag}(1/\boldsymbol{c})\boldsymbol{m}_{\widehat{\zeta}}$ is the mean vector of $\boldsymbol{D}\boldsymbol{V}^\intercal\,\mathbb{E}_q(\boldsymbol{\zeta}_1), \dots, \boldsymbol{D}\boldsymbol{V}^\intercal\,\mathbb{E}_q(\boldsymbol{\zeta}_n)$. Then,

$$
\begin{aligned}
&\sum_{i=1}^{n}(\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{m}_{\widehat{\zeta}})(\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{m}_{\widehat{\zeta}})^{\intercal}\\
&\qquad=\operatorname{diag}(\boldsymbol{c})\boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^{\intercal}\sum_{i=1}^{n}\left[\{\boldsymbol{D}_{\psi}\boldsymbol{V}_{\psi}^{\intercal}\,\mathbb{E}_q(\boldsymbol{\zeta}_i)-\boldsymbol{m}_{\zeta}\}\{\boldsymbol{D}_{\psi}\boldsymbol{V}_{\psi}^{\intercal}\,\mathbb{E}_q(\boldsymbol{\zeta}_i)-\boldsymbol{m}_{\zeta}\}^{\intercal}\right]\boldsymbol{Q}\boldsymbol{\Lambda}^{-1/2}\operatorname{diag}(\boldsymbol{c})\\
&\qquad=(n-1)\operatorname{diag}(\boldsymbol{c})\boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^{\intercal}\boldsymbol{C}_{\zeta}\boldsymbol{Q}\boldsymbol{\Lambda}^{-1/2}\operatorname{diag}(\boldsymbol{c})\\
&\qquad=(n-1)\operatorname{diag}(\boldsymbol{c})\boldsymbol{\Lambda}^{-1/2}\boldsymbol{Q}^{\intercal}\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\intercal}\boldsymbol{Q}\boldsymbol{\Lambda}^{-1/2}\operatorname{diag}(\boldsymbol{c})\\
&\qquad=(n-1)\operatorname{diag}(\boldsymbol{c}^2),
\end{aligned}
$$

which proves the results for the estimated scores.

Now we have

$$
\sum_{i=1}^{n}\widehat{y}_i(\boldsymbol{t}_g)=\sum_{i=1}^{n}\left\{\widehat{\mu}(\boldsymbol{t}_g)+\sum_{l=1}^{L}\widehat{\zeta}_{il}\widehat{\psi}_l(\boldsymbol{t}_g)\right\}=\sum_{i=1}^{n}\left\{\widehat{\mu}(\boldsymbol{t}_g)+\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\zeta}}_i\right\}=n\widehat{\mu}(\boldsymbol{t}_g)+n\widehat{\boldsymbol{\Psi}}\boldsymbol{m}_{\widehat{\zeta}},
$$

where $\widehat{\boldsymbol{\Psi}}\equiv\left[\,\widehat{\psi}_1(\boldsymbol{t}_g)\ \cdots\ \widehat{\psi}_L(\boldsymbol{t}_g)\,\right]$. Therefore, the sample covariance matrix of $\widehat{y}_1(\boldsymbol{t}_g),\ldots,\widehat{y}_n(\boldsymbol{t}_g)$ is such that

$$
\begin{aligned}
&\sum_{i=1}^{n}\left[\widehat{y}_i(\boldsymbol{t}_g)-\widehat{\mu}(\boldsymbol{t}_g)-\widehat{\boldsymbol{\Psi}}\boldsymbol{m}_{\widehat{\zeta}}\right]\left[\widehat{y}_i(\boldsymbol{t}_g)-\widehat{\mu}(\boldsymbol{t}_g)-\widehat{\boldsymbol{\Psi}}\boldsymbol{m}_{\widehat{\zeta}}\right]^{\intercal}\\
&\qquad=\sum_{i=1}^{n}\left(\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\zeta}}_i-\widehat{\boldsymbol{\Psi}}\boldsymbol{m}_{\widehat{\zeta}}\right)\left(\widehat{\boldsymbol{\Psi}}\widehat{\boldsymbol{\zeta}}_i-\widehat{\boldsymbol{\Psi}}\boldsymbol{m}_{\widehat{\zeta}}\right)^{\intercal}\\
&\qquad=\widehat{\boldsymbol{\Psi}}\left\{\sum_{i=1}^{n}(\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{m}_{\widehat{\zeta}})(\widehat{\boldsymbol{\zeta}}_i-\boldsymbol{m}_{\widehat{\zeta}})^{\intercal}\right\}\widehat{\boldsymbol{\Psi}}^{\intercal}\\
&\qquad=(n-1)\widehat{\boldsymbol{\Psi}}\operatorname{diag}(\boldsymbol{c}^2)\widehat{\boldsymbol{\Psi}}^{\intercal}.
\end{aligned}
$$

Simple rearrangement confirms that this is the eigenvalue decomposition of the sample covariance matrix of $\widehat{y}_1(\boldsymbol{t}_g),\ldots,\widehat{y}_n(\boldsymbol{t}_g)$, proving the result for the vectors $\widehat{\psi}_1(\boldsymbol{t}_g),\ldots,\widehat{\psi}_L(\boldsymbol{t}_g)$.

E Derivation of the Functional Principal Component Gaussian Likelihood Fragment

From (2.6), we have, for $i = 1, \dots, n$,

$$\log p(\boldsymbol{y}_i | \boldsymbol{\nu}, \boldsymbol{\zeta}_i, \sigma_\epsilon^2) = -\frac{T_i}{2} \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \left\| \boldsymbol{y}_i - \boldsymbol{C}_i \left(\boldsymbol{\nu}_\mu + \sum_{l=1}^{L} \zeta_{il} \boldsymbol{\nu}_{\psi_l} \right) \right\|^2 + \text{const.} \tag{E.1}$$

First, we establish the natural parameter vector for each of the optimal posterior density functions. These natural parameter vectors are essential for determining expectations with respect to the optimal posterior distribution. From equation (10) of Wand (2017), we deduce that the natural parameter vector for $q(\boldsymbol{\nu})$ is

$$\boldsymbol{\eta}_{q(\boldsymbol{\nu})} = \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2) \to \boldsymbol{\nu}} + \boldsymbol{\eta}_{\boldsymbol{\nu} \to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)},$$

the natural parameter vector for $q(\boldsymbol{\zeta}_i)$, $i = 1, \dots, n$, is

$$\boldsymbol{\eta}_{q(\boldsymbol{\zeta}_i)} = \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2) \to \boldsymbol{\zeta}_i} + \boldsymbol{\eta}_{\boldsymbol{\zeta}_i \to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)},$$

and the natural parameter vector for $q(\sigma_\epsilon^2)$ is

$$\boldsymbol{\eta}_{q(\sigma_\epsilon^2)} = \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2) \to \sigma_\epsilon^2} + \boldsymbol{\eta}_{\sigma_\epsilon^2 \to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}.$$

Next, we consider the updates for standard expectations that occur for each of the random variables and random vectors in (E.1). For $\boldsymbol{\nu}$, we need to determine the mean vector $\mathbb{E}_q(\boldsymbol{\nu})$ and the covariance matrix $\mathbb{C}\text{ov}_q(\boldsymbol{\nu})$. The expectations are taken with respect to the normalization of

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2) \to \boldsymbol{\nu}}(\boldsymbol{\nu}) m_{\boldsymbol{\nu} \to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}(\boldsymbol{\nu}),$$

which is a multivariate normal density function with natural parameter vector $\boldsymbol{\eta}_{q(\boldsymbol{\nu})}$. From (B.2), we have

$$\begin{aligned} \mathbb{E}_q(\boldsymbol{\nu}) &\longleftarrow -\frac{1}{2} \left[\text{vec}^{-1} \left\{ \left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})} \right)_2 \right\} \right]^{-1} \left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})} \right)_1 \\ \text{and} \quad \mathbb{C}\text{ov}_q(\boldsymbol{\nu}) &\longleftarrow -\frac{1}{2} \left[\text{vec}^{-1} \left\{ \left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})} \right)_2 \right\} \right]^{-1}. \end{aligned} \tag{E.2}$$

Furthermore, the mean vector has the form

$$\mathbb{E}_q(\boldsymbol{\nu}) \equiv \{\mathbb{E}_q(\boldsymbol{\nu}_\mu)^\intercal, \mathbb{E}_q(\boldsymbol{\nu}_{\psi_1})^\intercal, \dots, \mathbb{E}_q(\boldsymbol{\nu}_{\psi_L})^\intercal\}^\intercal, \tag{E.3}$$

and the covariance matrix has the form

$$\mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}) \equiv \begin{bmatrix} \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_\mu) & \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_\mu, \boldsymbol{\nu}_{\psi_1}) & \dots & \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_\mu, \boldsymbol{\nu}_{\psi_L}) \\ \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_{\psi_1}, \boldsymbol{\nu}_\mu) & \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_{\psi_1}) & \dots & \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_{\psi_1}, \boldsymbol{\nu}_{\psi_L}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_{\psi_L}, \boldsymbol{\nu}_\mu) & \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_{\psi_L}, \boldsymbol{\nu}_{\psi_1}) & \dots & \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_{\psi_L}) \end{bmatrix}. \tag{E.4}$$

Similarly, for each $i = 1, \dots, n$, we need to determine the optimal mean vector and covariance matrix for $\boldsymbol{\zeta}_i$, which are $\mathbb{E}_q(\boldsymbol{\zeta}_i)$ and $\mathbb{C}\mathrm{ov}_q(\boldsymbol{\zeta}_i)$, respectively. The expectations are taken with respect to the normalization of

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\zeta}_i}(\boldsymbol{\zeta}_i) m_{\boldsymbol{\zeta}_i\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}(\boldsymbol{\zeta}_i),$$

which is a multivariate normal density function with natural parameter vector $\boldsymbol{\eta}_{q(\boldsymbol{\zeta}_i)}$. According to (B.3),

$$\begin{aligned} &\mathbb{E}_q(\boldsymbol{\zeta}_i) \longleftarrow -\frac{1}{2}\left[\mathrm{vec}^{-1}\left\{\boldsymbol{D}_L^{+\intercal}\left(\boldsymbol{\eta}_{q(\boldsymbol{\zeta}_i)}\right)_2\right\}\right]^{-1}\left(\boldsymbol{\eta}_{q(\boldsymbol{\zeta}_i)}\right)_1 \\ \text{and}\quad &\mathbb{C}\mathrm{ov}_q(\boldsymbol{\zeta}_i) \longleftarrow -\frac{1}{2}\left[\mathrm{vec}^{-1}\left\{\boldsymbol{D}_L^{+\intercal}\left(\boldsymbol{\eta}_{q(\boldsymbol{\zeta}_i)}\right)_2\right\}\right]^{-1}, \quad \text{for } i = 1, \dots, n. \end{aligned} \tag{E.5}$$

Finally, for σ_ϵ^2, we need to determine $\mathbb{E}_q(1/\sigma_\epsilon^2)$, with the expectation taken with respect to the normalization of

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\sigma_\epsilon^2}(\sigma_\epsilon^2) m_{\sigma_\epsilon^2\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}(\sigma_\epsilon^2).$$

This is an inverse-χ^2 density function, with natural parameter vector $\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}$. According to Result 6 of Maestrini and Wand (2020),

$$\mathbb{E}_q(1/\sigma_\epsilon^2) \longleftarrow \frac{\left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_1 + 1}{\left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_2}.$$

Now, we turn our attention to the derivation of the message passed from $p(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_n, \sigma_\epsilon^2)$ to $\boldsymbol{\nu}$. Notice that

$$\boldsymbol{C}_i\left(\boldsymbol{\nu}_\mu - \sum_{l=1}^{L}\zeta_{il}\boldsymbol{\nu}_{\psi_l}\right) = (\widetilde{\boldsymbol{\zeta}}_i^{\intercal} \otimes \boldsymbol{C}_i)\boldsymbol{\nu}. \tag{E.6}$$

Therefore, as a function of $\boldsymbol{\nu}$, (E.1) can be re-written as

$$\begin{aligned}\log p(\boldsymbol{y}_i|\boldsymbol{\nu}, \boldsymbol{\zeta}_i, \sigma_\epsilon^2) &= -\frac{1}{2\sigma_\epsilon^2}\left|\left|\boldsymbol{y}_i - (\widetilde{\boldsymbol{\zeta}}_i^{\intercal} \otimes \boldsymbol{C}_i)\boldsymbol{\nu}\right|\right|^2 + \text{terms not involving } \boldsymbol{\nu} \\ &= \begin{bmatrix}\boldsymbol{\nu} \\ \text{vec}(\boldsymbol{\nu}\boldsymbol{\nu}^{\intercal})\end{bmatrix}^{\intercal}\begin{bmatrix}\frac{1}{\sigma_\epsilon^2}(\widetilde{\boldsymbol{\zeta}}_i^{\intercal} \otimes \boldsymbol{C}_i)^{\intercal}\boldsymbol{y}_i \\ -\frac{1}{2\sigma_\epsilon^2}\,\text{vec}\left\{(\widetilde{\boldsymbol{\zeta}}_i\widetilde{\boldsymbol{\zeta}}_i^{\intercal}) \otimes (\boldsymbol{C}_i^{\intercal}\boldsymbol{C}_i)\right\}\end{bmatrix} + \text{terms not involving } \boldsymbol{\nu}.\end{aligned}$$

According to equation (8) of Wand (2017), the message from the factor $p(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n, \sigma_\epsilon^2)$ to $\boldsymbol{\nu}$ is

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\nu}}(\boldsymbol{\nu}) \propto \exp\left\{\begin{bmatrix}\boldsymbol{\nu} \\ \text{vec}(\boldsymbol{\nu}\boldsymbol{\nu}^{\intercal})\end{bmatrix}^{\intercal}\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\nu}}\right\},$$

which is proportional to a multivariate normal density function. The update for the message's natural parameter vector, in (3.4), is dependent upon the mean vector and covariance matrix of $\widetilde{\boldsymbol{\zeta}}_i$, which are

$$\mathbb{E}_q(\widetilde{\zeta}_i) = \{1, \mathbb{E}_q(\boldsymbol{\zeta}_i)^{\intercal}\}^{\intercal} \quad \text{and} \quad \mathbb{C}\text{ov}_q(\widetilde{\zeta}_i) = \text{blockdiag}\left\{0, \mathbb{C}\text{ov}_q(\boldsymbol{\zeta}_i)\right\}, \quad \text{for } i = 1, \ldots, n, \tag{E.7}$$

where $\mathbb{E}_q(\boldsymbol{\zeta}_i)$ and $\mathbb{C}\text{ov}_q(\boldsymbol{\zeta}_i)$ are defined in (E.5). Note that a standard statistical result allows us to write

$$\mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_i\widetilde{\boldsymbol{\zeta}}_i^{\intercal}) = \mathbb{C}\text{ov}_q(\widetilde{\boldsymbol{\zeta}}_i) + \mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_i)\,\mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_i)^{\intercal}, \quad \text{for } i = 1, \ldots, n. \tag{E.8}$$

Next, notice that

$$\sum_{l=1}^{L}\zeta_{il}\boldsymbol{\nu}_{\psi_l} = \boldsymbol{V}_\psi\boldsymbol{\zeta}_i \tag{E.9}$$

Then, for each $i = 1, \ldots, n$, the log-density function in (E.1) can be represented as a function of $\boldsymbol{\zeta}_i$ by

$$
\begin{aligned}
\log p(\boldsymbol{y}_i|\boldsymbol{\nu},\boldsymbol{\zeta}_i,\sigma_\epsilon^2) &= -\frac{1}{2\sigma_\epsilon^2}\left|\left|\boldsymbol{y}_i - \boldsymbol{C}_i\boldsymbol{\nu}_\mu - \boldsymbol{C}_i\boldsymbol{V}_\psi\boldsymbol{\zeta}_i\right|\right|^2 + \text{terms not involving } \boldsymbol{\zeta}_i \\
&= \begin{bmatrix} \boldsymbol{\zeta}_i \\ \text{vech}(\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal) \end{bmatrix}^\intercal \begin{bmatrix} \frac{1}{\sigma_\epsilon^2}(\boldsymbol{V}_\psi^\intercal\boldsymbol{C}_i^\intercal\boldsymbol{y}_i - \boldsymbol{h}_{\mu\psi,i}) \\ -\frac{1}{2\sigma_\epsilon^2}\boldsymbol{D}_L^\intercal\,\text{vec}(\boldsymbol{H}_{\psi,i}) \end{bmatrix} + \text{terms not involving } \boldsymbol{\zeta}_i,
\end{aligned}
$$

According to equation (8) of Wand (2017), the message from the factor $p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)$ to $\boldsymbol{\zeta}_i$ is

$$
m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\zeta}_i}(\boldsymbol{\zeta}_i) \propto \exp\left\{\begin{bmatrix} \boldsymbol{\zeta}_i \\ \text{vech}(\boldsymbol{\zeta}_i\boldsymbol{\zeta}_i^\intercal) \end{bmatrix}^\intercal \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\zeta}_i}\right\},
$$

which is proportional to a multivariate normal density function. The message's natural parameter vector update, in (3.5), is dependant on the following expectations that are yet to be determined:

$$
\mathbb{E}_q(\boldsymbol{V}_\psi) \quad \text{and} \quad \mathbb{E}_q(\boldsymbol{H}_{\psi,i}), \quad \mathbb{E}_q(\boldsymbol{h}_{\mu\psi,i}), \quad i = 1,\ldots,n.
$$

Now, we have,

$$
\mathbb{E}_q(\boldsymbol{V}_\psi) = \begin{bmatrix} \mathbb{E}_q(\boldsymbol{\nu}_{\psi_1}) & \ldots & \mathbb{E}_q(\boldsymbol{\nu}_{\psi_L}) \end{bmatrix}, \tag{E.10}
$$

where, for $l = 1,\ldots,L$, $\mathbb{E}_q(\boldsymbol{\nu}_{\psi_l})$ is defined by (E.2) and (E.3). Next, $\mathbb{E}_q(\boldsymbol{h}_{\mu\psi,i})$ is an $L\times 1$ vector, with lth component being

$$
\mathbb{E}_q(\boldsymbol{h}_{\mu\psi,i})_l = \text{tr}\{\mathbb{C}\text{ov}_q(\boldsymbol{\nu}_\mu,\boldsymbol{\nu}_{\psi_l})\boldsymbol{C}_i^\intercal\boldsymbol{C}_i\} + \mathbb{E}_q(\boldsymbol{\nu}_{\psi_l})^\intercal\boldsymbol{C}_i^\intercal\boldsymbol{C}_i\,\mathbb{E}_q(\boldsymbol{\nu}_\mu), \quad l = 1,\ldots,L, \tag{E.11}
$$

which depends on sub-vectors of $\mathbb{E}_q(\boldsymbol{\nu})$ and sub-blocks of $\mathbb{C}\text{ov}_q(\boldsymbol{\nu})$ that are defined in (E.3) and (E.4), respectively. Finally, $\mathbb{E}_q(\boldsymbol{H}_{\psi,i})$ is an $L\times L$ matrix, with (l,l') component being

$$
\mathbb{E}_q(\boldsymbol{H}_{\psi,i})_{l,l'} = \text{tr}\{\mathbb{C}\text{ov}_q(\boldsymbol{\nu}_{\psi_{l'}},\boldsymbol{\nu}_{\psi_l})\boldsymbol{C}_i^\intercal\boldsymbol{C}_i\} + \mathbb{E}_q(\boldsymbol{\nu}_{\psi_l})^\intercal\boldsymbol{C}_i^\intercal\boldsymbol{C}_i\,\mathbb{E}_q(\boldsymbol{\nu}_{\psi_{l'}}), \quad l,l' = 1,\ldots,L. \tag{E.12}
$$

The final message to consider is the message from $p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)$ to σ_ϵ^2. As a function of σ_ϵ^2, (E.1) takes the form

$$\log p(\boldsymbol{y}_i|\boldsymbol{\nu}, \boldsymbol{\zeta}_i, \sigma_\epsilon^2) = -\frac{T_i}{2}\log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\left|\left|\boldsymbol{y}_i - \boldsymbol{C}_i\boldsymbol{V}\widetilde{\boldsymbol{\zeta}}_i\right|\right|^2 + \text{terms not involving } \sigma_\epsilon^2$$
$$= \begin{bmatrix}\log(\sigma_\epsilon^2)\\ \frac{1}{\sigma_\epsilon^2}\end{bmatrix}^\intercal \begin{bmatrix}-\frac{T_i}{2}\\ -\frac{1}{2}\left|\left|\boldsymbol{y}_i - \boldsymbol{C}_i\boldsymbol{V}\widetilde{\boldsymbol{\zeta}}_i\right|\right|^2\end{bmatrix} + \text{terms not involving } \sigma_\epsilon^2,$$

According to equation (8) of Wand (2017), the message from $p(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n, \sigma_\epsilon^2)$ to σ_ϵ^2 is

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\sigma_\epsilon^2}(\sigma_\epsilon^2) \propto \exp\left\{\begin{bmatrix}\log(\sigma_\epsilon^2)\\ 1/\sigma_\epsilon^2\end{bmatrix}^\intercal \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\sigma_\epsilon^2}\right\},$$

which is proportional to an inverse-χ^2 density function. The message's natural parameter vector, in (3.6), depends on the mean of the square norm $||\boldsymbol{y}_i - \boldsymbol{C}_i\boldsymbol{V}\widetilde{\boldsymbol{\zeta}}_i||^2$, for $i = 1, \ldots, n$. This expectation takes the form

$$\mathbb{E}_q\left(\left|\left|\boldsymbol{y}_i - \boldsymbol{C}_i\boldsymbol{V}\widetilde{\boldsymbol{\zeta}}_i\right|\right|^2\right) = \boldsymbol{y}_i^\intercal\boldsymbol{y}_i - 2\,\mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_i)^\intercal\,\mathbb{E}_q(\boldsymbol{V})^\intercal\boldsymbol{C}_i^\intercal\boldsymbol{y}_i$$
$$+ \operatorname{tr}\left[\left\{\mathbb{C}\text{ov}_q(\widetilde{\boldsymbol{\zeta}}_i) + \mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_i)\,\mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_i)^\intercal\right\}\mathbb{E}_q(\boldsymbol{H}_i)\right],$$

where we introduce the matrices

$$\boldsymbol{H}_i \equiv \begin{bmatrix}h_{\mu,i} & \boldsymbol{h}_{\mu\psi,i}^\intercal\\ \boldsymbol{h}_{\mu\psi,i} & \boldsymbol{H}_{\psi,i}\end{bmatrix}, \quad \text{for } i = 1, \ldots, n, \tag{E.13}$$

and vectors

$$h_{\mu,i} \equiv \boldsymbol{\nu}_\mu^\intercal\boldsymbol{C}_i^\intercal\boldsymbol{C}_i\boldsymbol{\nu}_\mu, \quad \text{for } i = 1, \ldots, n. \tag{E.14}$$

For each $i = 1, \ldots, n$, the mean vector $\mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_i)$ and $\mathbb{C}\text{ov}_q(\widetilde{\boldsymbol{\zeta}}_i)$ are defined in (E.7). However, $\mathbb{E}_q(\boldsymbol{V})$ and $\mathbb{E}_q(\boldsymbol{H}_i)$, $i = 1, \ldots, n$, are yet to be determined. We then have,

$$\mathbb{E}_q(\boldsymbol{V}) = \begin{bmatrix}\mathbb{E}_q(\boldsymbol{\nu}_\mu) & \mathbb{E}_q(\boldsymbol{\nu}_{\psi_1}) & \ldots & \mathbb{E}_q(\boldsymbol{\nu}_{\psi_L})\end{bmatrix},$$

where the component mean vectors are defined by (E.3). For each $i = 1, \ldots, n$, the expectation of $\boldsymbol{H}_i$, defined in (E.13), with respect to the optimal posterior distribution is

$$\mathbb{E}_q(\boldsymbol{H}_i) \equiv \begin{bmatrix} \mathbb{E}_q(h_{\mu,i}) & \mathbb{E}_q(\boldsymbol{h}_{\mu\psi,i})^\intercal \\ \mathbb{E}_q(\boldsymbol{h}_{\mu\psi,i}) & \mathbb{E}_q(\boldsymbol{H}_{\psi,i}) \end{bmatrix},$$

where $h_{\mu,i}$ is defined in (E.14) with expected value

$$\mathbb{E}_q(h_{\mu,i}) \equiv \operatorname{tr}\{\mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_\mu)\boldsymbol{C}_i^\intercal \boldsymbol{C}_i\} + \mathbb{E}_q(\boldsymbol{\nu}_\mu)^\intercal \boldsymbol{C}_i^\intercal \boldsymbol{C}_i \, \mathbb{E}_q(\boldsymbol{\nu}_\mu).$$

Furthermore, $\mathbb{E}_q(\boldsymbol{h}_{\mu\psi,i})$ and $\mathbb{E}_q(\boldsymbol{H}_{\psi,i})$ are defined in (E.11) and (E.12), respectively.

The FPCA Gaussian likelihood fragment, summarized in Algorithm 1, is a proceduralization of these results.

F Derivation of the Functional Principal Component Gaussian Penalization Fragment

From (2.6), we have, for $l = 1, \dots, L$,

$$\begin{aligned} \log p(\boldsymbol{\nu}_\mu, \boldsymbol{\nu}_{\psi_l} | \sigma^2_\mu, \sigma^2_{\psi_l}) = &- \frac{K}{2}\log(\sigma^2_\mu) - \frac{K}{2}\log(\sigma^2_{\psi_l}) - \frac{1}{2}(\boldsymbol{\beta}_\mu - \boldsymbol{\mu}_{\boldsymbol{\beta}_\mu})^\intercal \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}_\mu}(\boldsymbol{\beta}_\mu - \boldsymbol{\mu}_{\boldsymbol{\beta}_\mu}) \\ &- \frac{1}{2\sigma^2_\mu}\boldsymbol{u}_\mu^\intercal \boldsymbol{u}_\mu - \frac{1}{2}(\boldsymbol{\beta}_{\psi_l} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{\psi_l}})^\intercal \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\beta}_{\psi_l}}(\boldsymbol{\beta}_{\psi_l} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{\psi_l}}) - \frac{1}{2\sigma^2_{\psi_l}}\boldsymbol{u}_{\psi_l}^\intercal \boldsymbol{u}_{\psi_l}. \end{aligned} \tag{F.1}$$

First, we establish the natural parameter vector for each of the optimal posterior density functions. As explained in Appendix E, these natural parameter vectors are essential for determining expectations with respect to the optimal posterior distribution. According to equation (10) of Wand (2017), the natural parameter vector for $q(\boldsymbol{\nu})$ is

$$\boldsymbol{\eta}_{q(\boldsymbol{\nu})} = \boldsymbol{\eta}_{p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) \to \boldsymbol{\nu}} + \boldsymbol{\eta}_{\boldsymbol{\nu} \to p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L})},$$

the natural parameter vector for $q(\sigma^2_\mu)$ is

$$\boldsymbol{\eta}_{q(\sigma^2_\mu)} = \boldsymbol{\eta}_{p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) \to \sigma^2_\mu} + \boldsymbol{\eta}_{\sigma^2_\mu \to p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L})},$$

and, for $l = 1, \dots, L$, the natural parameter vector for $q(\sigma^2_{\psi_l})$ is

$$\boldsymbol{\eta}_{q(\sigma^2_{\psi_l})} = \boldsymbol{\eta}_{p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) \to \sigma^2_{\psi_l}} + \boldsymbol{\eta}_{\sigma^2_{\psi_l} \to p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L})}.$$

Next, we consider the updates for standard expectations of each of the random variables and random vectors that appear in (F.1). For $\boldsymbol{\nu}$, we require the mean vector $\mathbb{E}_q(\boldsymbol{\nu})$ and covariance matrix $\mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu})$ under the optimal posterior distribution. The expectations are taken with respect to the normalization of

$$m_{p(\boldsymbol{\nu}|\sigma^2_\mu,\sigma^2_{\psi_1},\dots,\sigma^2_{\psi_L})\to\boldsymbol{\nu}}(\boldsymbol{\nu})m_{\boldsymbol{\nu}\to p(\boldsymbol{\nu}|\sigma^2_\mu,\sigma^2_{\psi_1},\dots,\sigma^2_{\psi_L})}(\boldsymbol{\nu}),$$

which is a multivariate normal density function with natural parameter vector $\boldsymbol{\eta}_{q(\boldsymbol{\nu})}$. From (B.2), we have

$$\begin{aligned} \mathbb{E}_q(\boldsymbol{\nu}) &\longleftarrow -\frac{1}{2}\left[\mathrm{vec}^{-1}\left\{\left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})}\right)_2\right\}\right]^{-1}\left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})}\right)_1 \\ \text{and} \quad \mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}) &\longleftarrow -\frac{1}{2}\left[\mathrm{vec}^{-1}\left\{\left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})}\right)_2\right\}\right]^{-1}. \end{aligned} \tag{F.2}$$

The sub-vectors and sub-matrices of $\mathbb{E}_q(\boldsymbol{\nu})$ and $\mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu})$ are identical to those in (E.3) and (E.4), respectively. For the functional principal components Gaussian penalization fragment, however, we need to note further sub-vectors and sub-matrices. First,

$$\mathbb{E}_q(\boldsymbol{\nu}_\mu) \equiv \{\mathbb{E}_q(\boldsymbol{\beta}_\mu)^\intercal, \mathbb{E}_q(\boldsymbol{u}_\mu)^\intercal\}^\intercal \quad \text{and} \quad \mathbb{E}_q(\boldsymbol{\nu}_{\psi_l}) \equiv \{\mathbb{E}_q(\boldsymbol{\beta}_{\psi_l})^\intercal, \mathbb{E}_q(\boldsymbol{u}_{\psi_l})^\intercal\}^\intercal, \quad \text{for } l=1,\dots,L \tag{F.3}$$

and, second,

$$\mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_\mu) \equiv \begin{bmatrix} \mathbb{C}\mathrm{ov}_q(\boldsymbol{\beta}_\mu) & \mathbb{C}\mathrm{ov}_q(\boldsymbol{\beta}_\mu, \boldsymbol{u}_\mu) \\ \mathbb{C}\mathrm{ov}_q(\boldsymbol{u}_\mu, \boldsymbol{\beta}_\mu) & \mathbb{C}\mathrm{ov}_q(\boldsymbol{u}_\mu) \end{bmatrix} \tag{F.4}$$

and

$$\mathbb{C}\mathrm{ov}_q(\boldsymbol{\nu}_{\psi_l}) \equiv \begin{bmatrix} \mathbb{C}\mathrm{ov}_q(\boldsymbol{\beta}_{\psi_l}) & \mathbb{C}\mathrm{ov}_q(\boldsymbol{\beta}_{\psi_l}, \boldsymbol{u}_{\psi_l}) \\ \mathbb{C}\mathrm{ov}_q(\boldsymbol{u}_{\psi_l}, \boldsymbol{\beta}_{\psi_l}) & \mathbb{C}\mathrm{ov}_q(\boldsymbol{u}_{\psi_l}) \end{bmatrix}, \quad \text{for } l=1,\dots,L. \tag{F.5}$$

For σ^2_μ, we need to determine $\mathbb{E}_q(1/\sigma^2_\mu)$, with expectation taken with respect to the normalization of

$$m_{p(\boldsymbol{\nu}|\sigma^2_\mu,\sigma^2_{\psi_1},\dots,\sigma^2_{\psi_L})\to\sigma^2_\mu}(\sigma^2_\mu)m_{\sigma^2_\mu\to p(\boldsymbol{\nu}|\sigma^2_\mu,\sigma^2_{\psi_1},\dots,\sigma^2_{\psi_L})}(\sigma^2_\mu),$$

which is an inverse-χ^2 density function with natural parameter vector $\boldsymbol{\eta}_{q(\sigma_\mu^2)}$. According to Result 6 of Maestrini and Wand (2020),

$$\mathbb{E}_q(1/\sigma_\mu^2) \longleftarrow \frac{\left(\boldsymbol{\eta}_{q(\sigma_\mu^2)}\right)_1 + 1}{\left(\boldsymbol{\eta}_{q(\sigma_\mu^2)}\right)_2}. \tag{F.6}$$

Similar arguments can be used to show that

$$\mathbb{E}_q(1/\sigma_{\psi_l}^2) \longleftarrow \frac{\left(\boldsymbol{\eta}_{q(\sigma_{\psi_l}^2)}\right)_1 + 1}{\left(\boldsymbol{\eta}_{q(\sigma_{\psi_l}^2)}\right)_2}, \quad \text{for } l = 1, \dots, L. \tag{F.7}$$

Now, we turn our attention to the derivation of the messages passed from the factor. As a function of $\boldsymbol{\nu}$, (F.1) this can be re-written as

$$\begin{aligned}\log p(\boldsymbol{\nu}|\sigma_\mu^2, \sigma_{\psi_1}^2, \dots, \sigma_{\psi_L}^2) &= -\frac{1}{2}\boldsymbol{\nu}^\intercal \boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{\nu} + \boldsymbol{\nu}^\intercal \boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{\mu}_\nu + \text{terms not involving } \boldsymbol{\nu} \\ &= \begin{bmatrix} \boldsymbol{\nu} \\ \operatorname{vec}(\boldsymbol{\nu}\boldsymbol{\nu}^\intercal) \end{bmatrix}^\intercal \begin{bmatrix} \boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{\mu}_\nu \\ -\frac{1}{2}\operatorname{vec}(\boldsymbol{\Sigma}_\nu^{-1}) \end{bmatrix} + \text{terms not involving } \boldsymbol{\nu},\end{aligned}$$

According to equation (8) of Wand (2017), the message from the factor $p(\boldsymbol{\nu}|\sigma_\mu^2, \sigma_{\psi_1}^2, \dots, \sigma_{\psi_L}^2)$ to $\boldsymbol{\nu}$ is

$$m_{p(\boldsymbol{\nu}|\sigma_\mu^2,\sigma_{\psi_1}^2,\dots,\sigma_{\psi_L}^2)\to\boldsymbol{\nu}}(\boldsymbol{\nu}) \propto \exp\left\{ \begin{bmatrix} \boldsymbol{\nu} \\ \operatorname{vec}(\boldsymbol{\nu}\boldsymbol{\nu}^\intercal) \end{bmatrix}^\intercal \boldsymbol{\eta}_{p(\boldsymbol{\nu}|\sigma_\mu^2,\sigma_{\psi_1}^2,\dots,\sigma_{\psi_L}^2)\to\boldsymbol{\nu}} \right\},$$

which is proportional to a multivariate normal density function. The update for the message's natural parameter vector, in (3.7), is dependant upon the expectation of $\boldsymbol{\Sigma}_\nu^{-1}$, which is given by

$$\mathbb{E}_q(\boldsymbol{\Sigma}_\nu^{-1}) = \operatorname{blockdiag}\left\{ \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}_\mu} & \mathbf{O}^\intercal \\ \mathbf{O} & \mathbb{E}_q(1/\sigma_\mu^2)\boldsymbol{I}_K \end{bmatrix}, \operatorname*{blockdiag}_{l=1,\dots,L}\left(\begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\beta}_{\psi_l}} & \mathbf{O}^\intercal \\ \mathbf{O} & \mathbb{E}_q(1/\sigma_{\psi_l}^2)\boldsymbol{I}_K \end{bmatrix}\right)\right\},$$

where $\mathbb{E}_q(1/\sigma_\mu^2)$ and, for $l = 1, \dots, L$, $\mathbb{E}_q(1/\sigma_{\psi_l}^2)$ are defined in (F.6) and (F.7), respectively.

As a function of σ_μ^2, (F.1) can be re-written as

$$
\begin{aligned}
\log p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) &= -\frac{K}{2}\log(\sigma^2_\mu) - \frac{1}{2\sigma^2_\mu}\boldsymbol{u}_\mu^\intercal \boldsymbol{u}_\mu + \text{terms not involving } \sigma^2_\mu \\
&= \begin{bmatrix} \log(\sigma^2_\mu) \\ 1/\sigma^2_\mu \end{bmatrix}^\intercal \begin{bmatrix} -\frac{K}{2} \\ -\frac{1}{2}\boldsymbol{u}_\mu^\intercal \boldsymbol{u}_\mu \end{bmatrix} + \text{terms not involving } \sigma^2_\mu.
\end{aligned}
$$

According to equation (8) of Wand (2017), the message from the factor $p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L})$ to σ^2_μ is

$$
m_{p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) \to \sigma^2_\mu}(\sigma^2_\mu) \propto \exp\left\{ \begin{bmatrix} \log(\sigma^2_\mu) \\ 1/\sigma^2_\mu \end{bmatrix}^\intercal \boldsymbol{\eta}_{p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) \to \sigma^2_\mu} \right\},
$$

which is an inverse-χ^2 density function upon normalization. The message's natural parameter vector update in (3.8) depends on $\mathbb{E}_q(\boldsymbol{u}_\mu^\intercal \boldsymbol{u}_\mu)$. Standard statistical results and sub-vector and sub-matrix definitions in (F.3) and (F.4) can be employed to show that

$$
\mathbb{E}_q(\boldsymbol{u}_\mu^\intercal \boldsymbol{u}_\mu) = \mathbb{E}_q(\boldsymbol{u}_\mu)^\intercal \, \mathbb{E}_q(\boldsymbol{u}_\mu) + \operatorname{tr}\{\mathbb{C}\mathrm{ov}_q(\boldsymbol{u}_\mu)\}.
$$

As a function of $\sigma^2_{\psi_l}$, for $l = 1, \dots, L$, (F.1) can be re-written as

$$
\begin{aligned}
\log p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) &= -\frac{K}{2}\log(\sigma^2_{\psi_l}) - \frac{1}{2\sigma^2_{\psi_l}}\boldsymbol{u}_{\psi_l}^\intercal \boldsymbol{u}_{\psi_l} + \text{terms not involving } \sigma^2_{\psi_l} \\
&= \begin{bmatrix} \log(\sigma^2_{\psi_l}) \\ 1/\sigma^2_{\psi_l} \end{bmatrix}^\intercal \begin{bmatrix} -\frac{K}{2} \\ -\frac{1}{2}\boldsymbol{u}_{\psi_l}^\intercal \boldsymbol{u}_{\psi_l} \end{bmatrix} + \text{terms not involving } \sigma^2_{\psi_l}.
\end{aligned}
$$

According to equation (8) of Wand (2017), the message from the factor $p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L})$ to $\sigma^2_{\psi_l}$ is

$$
m_{p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) \to \sigma^2_\mu}(\sigma^2_{\psi_l}) \propto \exp\left\{ \begin{bmatrix} \log(\sigma^2_{\psi_l}) \\ 1/\sigma^2_{\psi_l} \end{bmatrix}^\intercal \boldsymbol{\eta}_{p(\boldsymbol{\nu}|\sigma^2_\mu, \sigma^2_{\psi_1}, \dots, \sigma^2_{\psi_L}) \to \sigma^2_{\psi_l}} \right\},
$$

which is an inverse-χ^2 density function upon normalization. The message's natural parameter vector update in (3.9) depends on $\mathbb{E}_q(\boldsymbol{u}_{\psi_l}^\intercal \boldsymbol{u}_{\psi_l})$. Standard statistical results and sub-vector and sub-matrix definitions in (F.3) and (F.5) can be employed to show that

$$\mathbb{E}_q(\boldsymbol{u}_{\psi_l}^\intercal \boldsymbol{u}_{\psi_l}) = \mathbb{E}_q(\boldsymbol{u}_{\psi_l})^\intercal \, \mathbb{E}_q(\boldsymbol{u}_{\psi_l}) + \text{tr}\left\{\mathbb{C}\text{ov}_q(\boldsymbol{u}_{\psi_l})\right\}.$$

The functional principal component Gaussian penalization fragment, summarized in Algorithm 2, is a proceduralization of these results.

G Two-level Sparse Matrix Background

Nolan and Wand (2020, Section 2) introduce two-level sparse matrix problems as a formal process for streamlining the the computations for two-level longitudinal data analysis (Pinheiro and Bates, 2000). Nolan et al. (2020, Section 4) extend these concepts to variational message passing for two-level linear mixed models. Here, we give a brief overview of these concepts.

A two-level sparse matrix problem is concerned with solving the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}$, where

$$\boldsymbol{A} \equiv \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12,1} & \boldsymbol{A}_{12,2} & \cdots & \boldsymbol{A}_{12,m} \\ \boldsymbol{A}_{12,1}^\intercal & \boldsymbol{A}_{22,1} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{A}_{12,2}^\intercal & \boldsymbol{O} & \boldsymbol{A}_{22,2} & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{12,m}^\intercal & \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{A}_{22,m} \end{bmatrix}, \quad \boldsymbol{a} \equiv \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_{2,1} \\ \vdots \\ \boldsymbol{a}_{2,m} \end{bmatrix} \quad \text{and} \quad \boldsymbol{x} \equiv \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_{2,1} \\ \vdots \\ \boldsymbol{x}_{2,m} \end{bmatrix} \tag{G.1}$$

and obtaining the submatrices of $\boldsymbol{A}^{-1}$ corresponding to the non-zero submatrices of $\boldsymbol{A}$:

$$\boldsymbol{A}^{-1} \equiv \begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12,1} & \boldsymbol{A}^{12,2} & \cdots & \boldsymbol{A}^{12,m} \\ \boldsymbol{A}^{12,1\intercal} & \boldsymbol{A}^{22,1} & \times & \cdots & \times \\ \boldsymbol{A}^{12,2\intercal} & \times & \boldsymbol{A}^{22,2} & \cdots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}^{12,m\intercal} & \times & \times & \cdots & \boldsymbol{A}^{22,m} \end{bmatrix}. \tag{G.2}$$

The sub-blocks of $\boldsymbol{A}^{-1}$ represented by $\times$ are not of interest because they correspond to between-group covariances in multilevel models. On the other hand, the sub-blocks of $\boldsymbol{A}^{-1}$ that are in the same position as the non-zero sub-blocks of $\boldsymbol{A}$ are required for obtaining standard errors of within-group fits.

Algorithm A.1 in the supplementary material of Nolan et al. (2020), called SolveTwoLevelSparseMatrix, outputs the partitioning of $\boldsymbol{x}_1$, $\boldsymbol{A}^{11}$ and $\{\boldsymbol{x}_{2,j}, \boldsymbol{A}^{12,j}, \boldsymbol{A}^{22,j}\}_{j=1,\dots,m}$ through streamlined computations that do not require direct inversion of the two-level sparse matrix $\boldsymbol{A}$.

H Derivation of the Multilevel Functional Principal Component Gaussian Likelihood Fragment

From (4.2), we have, for $i = 1, \dots, n$ and $j = 1, \dots, m_i$,

$$\begin{aligned}
&\log p(\boldsymbol{y}_{ij}|\boldsymbol{\nu}, \boldsymbol{\zeta}_i^{(1)}, \boldsymbol{\zeta}_{ij}^{(2)}, \sigma_\epsilon^2) = \\
&\quad -\frac{T_{ij}}{2}\log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\left|\left|\boldsymbol{y}_{ij} - \boldsymbol{C}_{ij}\left(\boldsymbol{\nu}_\mu + \sum_{l=1}^{L_1}\zeta_{il}^{(1)}\boldsymbol{\nu}_{\psi_l}^{(1)} + \sum_{l=1}^{L_2}\zeta_{ijl}^{(2)}\boldsymbol{\nu}_{\psi_l}^{(2)}\right)\right|\right|^2 \\
&\quad + \text{const.}
\end{aligned} \tag{H.1}$$

From equation (10) of Wand (2017), we deduce that the natural parameter vector for $q(\boldsymbol{\nu})$ is

$$\boldsymbol{\eta}_{q(\boldsymbol{\nu})} = \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\nu}} + \boldsymbol{\eta}_{\boldsymbol{\nu}\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)},$$

the natural parameter vector for $q(\boldsymbol{\zeta}_i)$, $i = 1, \dots, n$, is

$$\boldsymbol{\eta}_{q(\boldsymbol{\zeta}_i)} = \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\zeta}_i} + \boldsymbol{\eta}_{\boldsymbol{\zeta}_i\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)},$$

and the natural parameter vector for $q(\sigma_\epsilon^2)$ is

$$\boldsymbol{\eta}_{q(\sigma_\epsilon^2)} = \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\sigma_\epsilon^2} + \boldsymbol{\eta}_{\sigma_\epsilon^2\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}.$$

Next, we consider the updates for standard expectations that occur for each of the random variables and random vectors in (H.1). For $\boldsymbol{\nu}$, we need to determine the mean vector $\mathbb{E}_q(\boldsymbol{\nu})$ and the covariance matrix $\mathbb{C}\text{ov}_q(\boldsymbol{\nu})$. The expectations are taken with respect to the normalization of

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\nu}}(\boldsymbol{\nu})m_{\boldsymbol{\nu}\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}(\boldsymbol{\nu}),$$

which is a multivariate normal density function with natural parameter vector $\boldsymbol{\eta}_{q(\boldsymbol{\nu})}$. From (B.2), we have

$$\begin{aligned} \mathbb{E}_q(\boldsymbol{\nu}) &\longleftarrow -\frac{1}{2}\left[\text{vec}^{-1}\left\{\left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})}\right)_2\right\}\right]^{-1}\left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})}\right)_1 \\ \text{and}\quad \mathbb{C}\text{ov}_q(\boldsymbol{\nu}) &\longleftarrow -\frac{1}{2}\left[\text{vec}^{-1}\left\{\left(\boldsymbol{\eta}_{q(\boldsymbol{\nu})}\right)_2\right\}\right]^{-1}. \end{aligned} \tag{H.2}$$

In order to compute the q-density parameters for $q(\boldsymbol{\zeta}_i)$, note that the inverse covariance matrix has a two-level sparse structure as defined in (G.1). Naive computation of this matrix via methods similar to (E.5) requires inversion of this sparse two-level inverse covariance matrix. Instead, we turn to Corollary H.1, which is a direct consequence of Theorem 1 of Nolan and Wand (2020), for streamlined computations of these parameters. The solutions presented in Corollary H.1 can be computed through the function SolveTwoLevelSparseMatrix of Nolan et al. (2020, Algorithm A.1)

Corollary H.1. *For each $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$, The updates of the quantities $\mathbb{E}_q(\boldsymbol{\zeta}_i^{(1)})$, $\mathbb{E}_q(\boldsymbol{\zeta}_{ij}^{(2)})$, $\mathbb{C}ov_q\boldsymbol{\zeta}_i^{(1)}$, $\mathbb{C}ov_q(\boldsymbol{\zeta}_{ij}^{(2)})$ and $\mathbb{C}ov_q(\boldsymbol{\zeta}_i^{(1)}, \boldsymbol{\zeta}_{ij}^{(2)})$ with respect to the normalization of*

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\zeta}_i}(\boldsymbol{\zeta}_i)\; m_{\boldsymbol{\zeta}_i\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}(\boldsymbol{\zeta}_i),$$

are expressible as a two-level sparse matrix problem (see Appendix G) with

$$\boldsymbol{A} \equiv -2\begin{bmatrix} \text{vec}^{-1}(\boldsymbol{D}_{L_1}^\intercal \boldsymbol{\eta}_{1,2}) & \left[\frac{1}{2}\underset{j=1,\ldots,m_i}{\text{stack}}\{\text{vec}^{-1}(\boldsymbol{\eta}_{2,3,j})^\intercal\}\right]^\intercal \\ \frac{1}{2}\underset{j=1,\ldots,m_i}{\text{stack}}\{\text{vec}^{-1}(\boldsymbol{\eta}_{2,3,j})^\intercal\} & \underset{j=1,\ldots,m_i}{\text{blockdiag}}\left\{\text{vec}^{-1}(\boldsymbol{D}_{L_2}^{+\intercal}\boldsymbol{\eta}_{2,2,j})\right\} \end{bmatrix}$$

and

$$\boldsymbol{a} \equiv \begin{bmatrix} \boldsymbol{\eta}_{1,1} \\ \underset{j=1,\ldots,m_i}{\text{stack}}(\boldsymbol{\eta}_{2,1,j}) \end{bmatrix}, \quad \textit{where} \quad \begin{bmatrix} \boldsymbol{\eta}_{1,1}\ (L_1 \times 1) \\ \boldsymbol{\eta}_{1,2}\ (\frac{1}{2}L_1(L_1+1)\times 1) \\ \underset{j=1,\ldots,m_i}{\text{stack}}\left(\begin{bmatrix} \boldsymbol{\eta}_{2,1,j}\ (L_2\times 1) \\ \boldsymbol{\eta}_{2,2,j}\ (\frac{1}{2}L_2(L_2+1)\times 1) \\ \boldsymbol{\eta}_{2,3,j}\ (L_1L_2\times 1)\end{bmatrix}\right) \end{bmatrix}$$

is the partitioning of $\boldsymbol{\eta}_{q(\boldsymbol{\zeta}_i)}$ that defines $\boldsymbol{\eta}_{1,1}$, $\boldsymbol{\eta}_{1,2}$ and $\{\boldsymbol{\eta}_{2,1,j}, \boldsymbol{\eta}_{2,2,j}, \boldsymbol{\eta}_{2,3,j}\}_{j=1,\ldots,m_i}$. The solutions, according to (G.1) and (G.2), are $\mathbb{E}_q(\boldsymbol{\zeta}_i^{(1)}) = \boldsymbol{x}_1$, $\mathbb{C}ov_q(\boldsymbol{\zeta}_i^{(1)}) = \boldsymbol{A}^{11}$, $\mathbb{E}_q(\boldsymbol{\zeta}_{ij}^{(2)}) = \boldsymbol{x}_{2,j}$, $\mathbb{C}ov_q(\boldsymbol{\zeta}_{ij}^{(2)}) = \boldsymbol{A}^{22,j}$ and $\mathbb{C}ov_q(\boldsymbol{\zeta}_i^{(1)}, \boldsymbol{\zeta}_{ij}^{(2)}) = \boldsymbol{A}^{12,j}$.

Proof. Note that

$$
\begin{aligned}
q(\boldsymbol{\zeta}_i) &\propto m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\zeta}_i}(\boldsymbol{\zeta}_i)\; m_{\boldsymbol{\zeta}_i\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}(\boldsymbol{\zeta}_i) \\
&= \exp\left\{ \begin{bmatrix} \boldsymbol{\zeta}_i^{(1)} \\ \mathrm{vech}(\boldsymbol{\zeta}_i^{(1)}\boldsymbol{\zeta}_i^{(1)\intercal}) \\ \underset{j=1,\ldots,m_i}{\mathrm{stack}}\left\{ \begin{bmatrix} \boldsymbol{\zeta}_{ij}^{(2)} \\ \mathrm{vech}(\boldsymbol{\zeta}_{ij}^{(2)}\boldsymbol{\zeta}_{ij}^{(2)\intercal}) \\ \mathrm{vec}(\boldsymbol{\zeta}_i^{(1)}\boldsymbol{\zeta}_{ij}^{(2)\intercal}) \end{bmatrix} \right\} \end{bmatrix}^\intercal \boldsymbol{\eta}_{q(\boldsymbol{\zeta}_i)} \right\} \\
&= \exp\left(\boldsymbol{\zeta}_i^\intercal \boldsymbol{a} - \frac{1}{2}\boldsymbol{\zeta}_i^\intercal \boldsymbol{A}\boldsymbol{\zeta}_i \right).
\end{aligned}
$$

Standard manipulations then lead to

$$\mathbb{E}_q(\boldsymbol{\zeta}_i) = \boldsymbol{A}^{-1}\boldsymbol{a} \quad \text{and} \quad \mathbb{C}\mathrm{ov}_q(\boldsymbol{\zeta}_i) = \boldsymbol{A}^{-1}$$

The solutions then follow from extraction of the sub-vectors of $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{a}$ and the important sub-blocks of $\boldsymbol{A}^{-1}$ according to (G.2). □

Finally, for σ_ϵ^2, we need to determine $\mathbb{E}_q(1/\sigma_\epsilon^2)$, with the expectation taken with respect to the normalization of

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\sigma_\epsilon^2}(\sigma_\epsilon^2)\, m_{\sigma_\epsilon^2\to p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)}(\sigma_\epsilon^2).$$

This is an inverse-χ^2 density function, with natural parameter vector $\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}$. According to Result 6 of Maestrini and Wand (2020),

$$\mathbb{E}_q(1/\sigma_\epsilon^2) \longleftarrow \frac{\left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_1 + 1}{\left(\boldsymbol{\eta}_{q(\sigma_\epsilon^2)}\right)_2}.$$

Now, we turn our attention to the derivation of the message passed from $p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)$ to $\boldsymbol{\nu}$. Notice that

$$\boldsymbol{C}_{ij}\left(\boldsymbol{\nu}_\mu + \sum_{l=1}^{L_1} \zeta_{il}^{(1)}\boldsymbol{\nu}_{\psi_l}^{(1)} + \sum_{l=1}^{L_2} \zeta_{ijl}^{(2)}\boldsymbol{\nu}_{\psi_l}^{(2)} \right) = (\widetilde{\boldsymbol{\zeta}}_{ij}^\intercal \otimes \boldsymbol{C}_{ij})\boldsymbol{\nu}. \tag{H.3}$$

Therefore, as a function of $\boldsymbol{\nu}$, (H.1) can be re-written as

$$
\begin{aligned}
&\log p(\boldsymbol{y}_{ij}|\boldsymbol{\nu}, \boldsymbol{\zeta}_i^{(1)}, \boldsymbol{\zeta}_{ij}^{(2)}, \sigma_\epsilon^2) \\
&\quad = -\frac{1}{2\sigma_\epsilon^2}\left|\left|\boldsymbol{y}_{ij} - (\widetilde{\boldsymbol{\zeta}}_{ij}^\intercal \otimes \boldsymbol{C}_{ij})\boldsymbol{\nu}\right|\right|^2 + \text{terms not involving } \boldsymbol{\nu} \\
&\quad = \begin{bmatrix} \boldsymbol{\nu} \\ \text{vec}(\boldsymbol{\nu}\boldsymbol{\nu}^\intercal) \end{bmatrix}^\intercal \begin{bmatrix} \frac{1}{\sigma_\epsilon^2}(\widetilde{\boldsymbol{\zeta}}_{ij}^\intercal \otimes \boldsymbol{C}_{ij})^\intercal \boldsymbol{y}_{ij} \\ -\frac{1}{2\sigma_\epsilon^2}\text{vec}\left\{(\widetilde{\boldsymbol{\zeta}}_i\widetilde{\boldsymbol{\zeta}}_i^\intercal) \otimes (\boldsymbol{C}_i^\intercal \boldsymbol{C}_i)\right\} \end{bmatrix} + \text{terms not involving } \boldsymbol{\nu}.
\end{aligned}
$$

According to equation (8) of Wand (2017), the message from the factor $p(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n, \sigma_\epsilon^2)$ to $\boldsymbol{\nu}$ is

$$
m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\nu}}(\boldsymbol{\nu}) \propto \exp\left\{\begin{bmatrix} \boldsymbol{\nu} \\ \text{vec}(\boldsymbol{\nu}\boldsymbol{\nu}^\intercal) \end{bmatrix}^\intercal \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\nu}}\right\},
$$

which is proportional to a multivariate normal density function. The update for the message's natural parameter vector, in (4.4), is dependent upon the mean vector and covariance matrix of $\widetilde{\boldsymbol{\zeta}}_i$, which are

$$
\mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_{ij}) = \begin{bmatrix} 1 \\ \mathbb{E}_q(\boldsymbol{\zeta}_i^{(1)}) \\ \mathbb{E}_q(\boldsymbol{\zeta}_{ij}^{(2)}) \end{bmatrix} \quad \text{and} \quad \mathbb{C}\text{ov}_q(\widetilde{\boldsymbol{\zeta}}_{ij}) = \begin{bmatrix} 0 & \mathbf{0}_{L_1}^\intercal & \mathbf{0}_{L_2}^\intercal \\ \mathbf{0}_{L_1} & \mathbb{C}\text{ov}_q(\boldsymbol{\zeta}_i^{(1)}) & \mathbb{C}\text{ov}_q(\boldsymbol{\zeta}_i^{(1)}, \boldsymbol{\zeta}_{ij}^{(2)}) \\ \mathbf{0}_{L_2} & \mathbb{C}\text{ov}_q(\boldsymbol{\zeta}_{ij}^{(2)}, \boldsymbol{\zeta}_i^{(1)}) & \mathbb{C}\text{ov}_q(\boldsymbol{\zeta}_{ij}^{(2)}) \end{bmatrix}, \tag{H.4}
$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. Note that a standard statistical result allows us to write

$$
\mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_{ij}\widetilde{\boldsymbol{\zeta}}_{ij}^\intercal) = \mathbb{C}\text{ov}_q(\widetilde{\boldsymbol{\zeta}}_{ij}) + \mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_{ij})\,\mathbb{E}_q(\widetilde{\boldsymbol{\zeta}}_{ij})^\intercal. \tag{H.5}
$$

Next, notice that

$$
\sum_{l=1}^{L_1} \zeta_{il}^{(1)} \boldsymbol{\nu}_{\psi_l}^{(1)} = \boldsymbol{V}_\psi^{(1)} \boldsymbol{\zeta}_i^{(1)} \quad \text{and} \quad \sum_{l=1}^{L_2} \zeta_{ijl}^{(2)} \boldsymbol{\nu}_{\psi_l}^{(2)} = \boldsymbol{V}_\psi^{(2)} \boldsymbol{\zeta}_{ij}^{(2)}. \tag{H.6}
$$

Then, for each $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$, the log-density function in (H.1) can be represented as a function of $\boldsymbol{\zeta}_i$ by

$$
\begin{aligned}
&\log p(\boldsymbol{y}_i|\boldsymbol{\nu},\boldsymbol{\zeta}_i,\sigma_\epsilon^2)\\
&\quad=\sum_{j=1}^{m_i}\log p(\boldsymbol{y}_{ij}|\boldsymbol{\nu},\boldsymbol{\zeta}_i^{(1)},\boldsymbol{\zeta}_{ij}^{(2)},\sigma_\epsilon^2)\\
&\quad=-\frac{1}{2\sigma_\epsilon^2}\sum_{j=1}^{m_i}\left|\left|\boldsymbol{y}_{ij}-\boldsymbol{C}_{ij}(\boldsymbol{\nu}_\mu+\boldsymbol{V}_\psi^{(1)}\boldsymbol{\zeta}_i^{(1)}+\boldsymbol{V}_\psi^{(2)}\boldsymbol{\zeta}_{ij}^{(2)})\right|\right|^2+\text{terms not involving }\boldsymbol{\zeta}_i\\
&\quad=\begin{bmatrix}\boldsymbol{\zeta}_i^{(1)}\\ \mathrm{vech}(\boldsymbol{\zeta}_i^{(1)}\boldsymbol{\zeta}_i^{(1)\intercal})\\ \underset{j=1,\dots,m_i}{\mathrm{stack}}\left\{\begin{bmatrix}\boldsymbol{\zeta}_{ij}^{(2)}\\ \mathrm{vech}(\boldsymbol{\zeta}_{ij}^{(2)}\boldsymbol{\zeta}_{ij}^{(2)\intercal})\\ \mathrm{vec}(\boldsymbol{\zeta}_i^{(1)}\boldsymbol{\zeta}_{ij}^{(2)\intercal})\end{bmatrix}\right\}\end{bmatrix}^\intercal\begin{bmatrix}\frac{1}{\sigma_\epsilon^2}\sum_{j=1}^{m_i}(\boldsymbol{V}_\psi^{(1)\intercal}\boldsymbol{C}_{ij}^\intercal\boldsymbol{y}_{ij}-\boldsymbol{h}_{\mu\psi,ij}^{(1)})\\ -\frac{1}{2\sigma_\epsilon^2}\boldsymbol{D}_{L_1}^\intercal\sum_{j=1}^{m_i}\mathrm{vec}(\boldsymbol{H}_{\psi,ij}^{(1,1)})\\ \underset{j=1,\dots,m_i}{\mathrm{stack}}\left\{\begin{bmatrix}\frac{1}{\sigma_\epsilon^2}(\boldsymbol{V}_\psi^{(2)\intercal}\boldsymbol{C}_{ij}^\intercal\boldsymbol{y}_{ij}-\boldsymbol{h}_{\mu\psi,ij}^{(2)})\\ -\frac{1}{2\sigma_\epsilon^2}\boldsymbol{D}_{L_2}^\intercal\mathrm{vec}(\boldsymbol{H}_{\psi,ij}^{(2,2)})\\ -\frac{1}{\sigma_\epsilon^2}\mathrm{vec}(\boldsymbol{H}_{\psi,ij}^{(1,2)})\end{bmatrix}\right\}\end{bmatrix}\\
&\qquad\qquad+\text{terms not involving }\boldsymbol{\zeta}_i,
\end{aligned}
$$

According to equation (8) of Wand (2017), the message from the factor $p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)$ to $\boldsymbol{\zeta}_i$ is

$$
m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\zeta}_i}(\boldsymbol{\zeta}_i)\propto\exp\left\{\begin{bmatrix}\boldsymbol{\zeta}_i^{(1)}\\ \mathrm{vech}(\boldsymbol{\zeta}_i^{(1)}\boldsymbol{\zeta}_i^{(1)\intercal})\\ \underset{j=1,\dots,m_i}{\mathrm{stack}}\left\{\begin{bmatrix}\boldsymbol{\zeta}_{ij}^{(2)}\\ \mathrm{vech}(\boldsymbol{\zeta}_{ij}^{(2)}\boldsymbol{\zeta}_{ij}^{(2)\intercal})\\ \mathrm{vec}(\boldsymbol{\zeta}_i^{(1)}\boldsymbol{\zeta}_{ij}^{(2)\intercal})\end{bmatrix}\right\}\end{bmatrix}^\intercal\boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\dots,\boldsymbol{\zeta}_n,\sigma_\epsilon^2)\to\boldsymbol{\zeta}_i}\right\},
$$

which is proportional to a multivariate normal density function. The message's natural parameter vector update, in (4.6), is dependant on the following expectations that are yet to be determined:

$$
\mathbb{E}_q(\boldsymbol{V}_\psi^{(r)}),\quad\mathbb{E}_q(\boldsymbol{h}_{\mu\psi,ij}^{(r)})\quad\text{and}\quad\mathbb{E}_q(\boldsymbol{H}_{\psi,ij}^{(r,s)}),
$$

for $r,s=1,2$, $i=1,\dots n$ and $j=1,\dots,m_i$. Now, we have,

$$
\mathbb{E}_q(\boldsymbol{V}_\psi^{(r)})=\left[\mathbb{E}_q(\boldsymbol{\nu}_{\psi_1}^{(r)})\quad\dots\quad\mathbb{E}_q(\boldsymbol{\nu}_{\psi_{L_r}}^{(r)})\right].\tag{H.7}
$$

Next, $\mathbb{E}_q(\boldsymbol{h}_{\mu\psi,ij}^{(r)})$ is an $L_r\times 1$ vector, with lth component being

$$\mathbb{E}_q(\boldsymbol{h}^{(r)}_{\mu\psi,ij})_l = \text{tr}\{\mathbb{C}\text{ov}_q(\boldsymbol{\nu}_\mu, \boldsymbol{\nu}^{(r)}_{\psi_l})\boldsymbol{C}^\intercal_{ij}\boldsymbol{C}_{ij}\} + \mathbb{E}_q(\boldsymbol{\nu}^{(r)}_{\psi_l})^\intercal \boldsymbol{C}^\intercal_{ij}\boldsymbol{C}_{ij}\,\mathbb{E}_q(\boldsymbol{\nu}_\mu), \quad \text{(H.8)}$$

for $l = 1, \ldots, L_r$. Finally, $\mathbb{E}_q(\boldsymbol{H}^{(r,s)}_{\psi,ij})$ is an $L_r \times L_s$ matrix, with (l, l') component being

$$\mathbb{E}_q(\boldsymbol{H}^{(r,s)}_{\psi,ij})_{l,l'} = \text{tr}\{\mathbb{C}\text{ov}_q(\boldsymbol{\nu}^{(s)}_{\psi_{l'}}, \boldsymbol{\nu}^{(r)}_{\psi_l})\boldsymbol{C}^\intercal_{ij}\boldsymbol{C}_{ij}\} + \mathbb{E}_q(\boldsymbol{\nu}^{(r)}_{\psi_l})^\intercal \boldsymbol{C}^\intercal_{ij}\boldsymbol{C}_{ij}\,\mathbb{E}_q(\boldsymbol{\nu}^{(s)}_{\psi_{l'}}), \quad \text{(H.9)}$$

for $l = 1, \ldots, L_r$ and $l' = 1, \ldots, L_s$.

The final message to consider is the message from $p(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n, \sigma^2_\epsilon)$ to σ^2_ϵ. As a function of σ^2_ϵ, (H.1) takes the form

$$\begin{aligned}
\log p(\boldsymbol{y}_{ij}|\boldsymbol{\nu}, \boldsymbol{\zeta}^{(1)}_i, \boldsymbol{\zeta}^{(2)}_{ij}, \sigma^2_\epsilon) &= -\frac{T_{ij}}{2}\log(\sigma^2_\epsilon) - \frac{1}{2\sigma^2_\epsilon}\left|\left|\boldsymbol{y}_{ij} - \boldsymbol{C}_{ij}(\boldsymbol{\nu}_\mu - \boldsymbol{V}^{(1)}_\psi\boldsymbol{\zeta}^{(1)}_i - \boldsymbol{V}^{(2)}_\psi\boldsymbol{\zeta}^{(2)}_{ij})\right|\right|^2 \\
&\quad + \text{terms not involving } \sigma^2_\epsilon \\
&= \begin{bmatrix} \log(\sigma^2_\epsilon) \\ \frac{1}{\sigma^2_\epsilon} \end{bmatrix}^\intercal \begin{bmatrix} -\frac{T_{ij}}{2} \\ -\frac{1}{2}\left|\left|\boldsymbol{y}_{ij} - \boldsymbol{C}_{ij}(\boldsymbol{\nu}_\mu - \boldsymbol{V}^{(1)}_\psi\boldsymbol{\zeta}^{(1)}_i - \boldsymbol{V}^{(2)}_\psi\boldsymbol{\zeta}^{(2)}_{ij})\right|\right|^2 \end{bmatrix} \\
&\quad + \text{terms not involving } \sigma^2_\epsilon.
\end{aligned}$$

According to equation (8) of Wand (2017), the message from $p(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n, \sigma^2_\epsilon)$ to σ^2_ϵ is

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma^2_\epsilon)\to\sigma^2_\epsilon}(\sigma^2_\epsilon) \propto \exp\left\{\begin{bmatrix} \log(\sigma^2_\epsilon) \\ 1/\sigma^2_\epsilon \end{bmatrix}^\intercal \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}_1,\ldots,\boldsymbol{\zeta}_n,\sigma^2_\epsilon)\to\sigma^2_\epsilon}\right\},$$

which is proportional to an inverse-χ^2 density function. The message's natural parameter vector, in (4.7), depends on the mean of the square norm $||\boldsymbol{y}_{ij} - \boldsymbol{C}_{ij}(\boldsymbol{\nu}_\mu - \boldsymbol{V}^{(1)}_\psi\boldsymbol{\zeta}^{(1)}_i - \boldsymbol{V}^{(2)}_\psi\boldsymbol{\zeta}^{(2)}_{ij})||^2$, for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. Setting $h_{\mu,ij} \equiv \boldsymbol{\nu}^\intercal_\mu\boldsymbol{C}^\intercal_{ij}\boldsymbol{C}_{ij}\boldsymbol{\nu}_\mu$, with q-expectation

$$\mathbb{E}_q(h_{\mu,ij}) = \text{tr}\{\mathbb{C}\text{ov}_q(\boldsymbol{\nu}_\mu)\boldsymbol{C}^\intercal_{ij}\boldsymbol{C}_{ij}\} + \mathbb{E}_q(\boldsymbol{\nu}_\mu)^\intercal\boldsymbol{C}^\intercal_{ij}\boldsymbol{C}_{ij}\,\mathbb{E}_q(\boldsymbol{\nu}_\mu).$$

Streamlined computations for this expectation are such that

$$
\mathbb{E}_q\left\{\left|\left|\boldsymbol{y}_{ij}-\boldsymbol{C}_{ij}(\boldsymbol{\nu}_\mu-\boldsymbol{V}_\psi^{(1)}\boldsymbol{\zeta}_i^{(1)}-\boldsymbol{V}_\psi^{(2)}\boldsymbol{\zeta}_{ij}^{(2)})\right|\right|^2\right\}=
$$

$$
\begin{aligned}
&\boldsymbol{y}_{ij}^\intercal\boldsymbol{y}_{ij}-2\,\mathbb{E}_q(\boldsymbol{\nu}_\mu)^\intercal\boldsymbol{C}_{ij}^\intercal\boldsymbol{y}_{ij}-2\,\mathbb{E}_q(\boldsymbol{\zeta}_i^{(1)})^\intercal\,\mathbb{E}_q(\boldsymbol{V}_\psi^{(1)})^\intercal\boldsymbol{C}_{ij}^\intercal\boldsymbol{y}_{ij}\\
&\quad-2\,\mathbb{E}_q(\boldsymbol{\zeta}_{ij}^{(2)})^\intercal\,\mathbb{E}_q(\boldsymbol{V}_\psi^{(2)})^\intercal\boldsymbol{C}_{ij}^\intercal\boldsymbol{y}_{ij}+\mathbb{E}_q(h_{\mu,ij})+2\,\mathbb{E}_q(\boldsymbol{\zeta}_i^{(1)})^\intercal\,\mathbb{E}_q(\boldsymbol{h}_{\mu\psi,ij}^{(1)})\\
&\quad+2\,\mathbb{E}_q(\boldsymbol{\zeta}_{ij}^{(2)})^\intercal\,\mathbb{E}_q(\boldsymbol{h}_{\mu\psi,ij}^{(2)})+\mathrm{tr}\{\mathbb{E}_q(\boldsymbol{\zeta}_i^{(1)}\boldsymbol{\zeta}_i^{(1)\intercal})\,\mathbb{E}_q(\boldsymbol{H}_{\psi,ij}^{(1,1)})\}\\
&\quad+2\,\mathrm{tr}\{\mathbb{E}_q(\boldsymbol{H}_{\psi,ij}^{(1,2)})\,\mathbb{E}_q(\boldsymbol{\zeta}_{ij}^{(2)}\boldsymbol{\zeta}_i^{(1)\intercal})\}+\mathrm{tr}\{\mathbb{E}_q(\boldsymbol{\zeta}_{ij}^{(2)}\boldsymbol{\zeta}_{ij}^{(2)\intercal})\,\mathbb{E}_q(\boldsymbol{H}_{\psi,ij}^{(2,2)})\}.
\end{aligned}
$$

The FPCA Gaussian likelihood fragment, summarized in Algorithm 3, is a proceduralization of these results.

I Convergence and Algorithmic Updates

I.1 Convergence

Variational Bayesian inference, and hence VMP, is based on the notion of minimal Kullback-Leibler divergence to approximate a posterior density function. For arbitrary density functions p_1 and p_2 on $\mathbb{R}^d$, the Kullback-Leibler divergence of p_1 from p_2 is

$$
D_{\mathrm{KL}}(p_1,p_2)\equiv\int\mathbb{R}^d\log\left\{\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\right\}p_1(\boldsymbol{x})d\boldsymbol{x}.
$$

Note that

$$
D_{\mathrm{KL}}(p_1,p_2)\geq 0. \tag{I.1}
$$

Consider a generic Bayesian model with observed data vector $\boldsymbol{y}$ and parameter vector $\boldsymbol{\theta}\in\Theta$, where $\boldsymbol{\Theta}$ is a parameter space. We make the assumption that $\boldsymbol{y}$ and $\boldsymbol{\theta}$ are continuous random variables with density functions $p(\boldsymbol{y})$ and $p(\boldsymbol{\theta})$. For the case where some components are discrete, a similar treatment applies with summations replacing integrals. Next, let $q(\boldsymbol{\theta})$ represent an arbitrary density function over the parameter space Θ. The essence of variational Bayesian inference is to restrict q to a class of density functions $\mathcal{Q}$ and use the optimal q-density function, defined by

$$
q^*(\boldsymbol{\theta})\equiv\underset{q\in\mathcal{Q}}{\mathrm{argmin}}\;D_{\mathrm{KL}}\{q(\boldsymbol{\theta}),p(\boldsymbol{\theta}|\boldsymbol{y})\}, \tag{I.2}
$$

as an approximation to the true posterior density function $p(\boldsymbol{\theta}|\boldsymbol{y})$.

Simple algebraic arguments (Ormerod and Wand, 2010, Section 2.1) show that the marginal log-likelihood satisfies:

$$\log p(\boldsymbol{y}) = D_{\mathrm{KL}}\{q(\boldsymbol{\theta}), p(\boldsymbol{\theta}|\boldsymbol{y})\} + \log \underline{p}(\boldsymbol{y}; q),$$

where

$$\underline{p}(\boldsymbol{y}; q) \equiv \exp\left[\int \log\left\{\frac{p(\boldsymbol{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right\} q(\boldsymbol{\theta}) d\boldsymbol{\theta}\right],$$

implying that

$$\log \underline{p}(\boldsymbol{y}; q) = \mathbb{E}_q\{\log p(\boldsymbol{y}, \boldsymbol{\theta})\} - \mathbb{E}_q\{\log q(\boldsymbol{\theta})\}.$$

From the non-negativity condition of (I.1), we have

$$\log \underline{p}(\boldsymbol{y}; q) \leq \log p(\boldsymbol{y})$$

showing that $\log \underline{p}(\boldsymbol{y}; q)$ is a lower-bound on the marginal log-likelihood. This leads to an equivalent form for the optimisation problem in (I.2):

$$q^*(\boldsymbol{\theta}) \equiv \underset{q \in \mathcal{Q}}{\operatorname{argmax}}\{\log \underline{p}(\boldsymbol{y}; q)\}. \tag{I.3}$$

This alternate expression has the advantage of representing the optimal q-density function as maximising the lower-bound on the marginal log-likelihood Rohde and Wand (2016). For the remainder of this article, we will address variational Bayesian inference with (I.3), rather than (I.2).

I.2 VMP Algorithms

Once the functional forms of the messages have been determined, the VMP iteration loop has the following generic steps:

- Choose a fragment.
- Update the parameter vectors of the messages passed from the factor's neighboring stochastic nodes to the factor.
- Update the parameter vectors of the messages passed from the factor to its neighboring stochastic nodes.

Pseudocode for the VMP algorithm for the Bayesian FPCA model in (2.6) is provided in Algorithm 1. Likewise, pseudocode for the VMP algorithm for the Bayesian MlFPCA model in (4.2) is provided in Algorithm 2

Algorithm 1 Generic VMP algorithm for the Gaussian response FPCA model (2.6) with mean field restriction (3.1).

Inputs: All hyperparameters and observed data
Initialize: All factor to stochastic node messages. ▷ Wand (2017, Section 2.5)
Updates:
while log $\underline{p}(\boldsymbol{y}; q)$ has not converged **do**
  Update all stochastic node to factor messages. ▷ Wand (2017, Section 2.5)
  Update the fragment for $p(\boldsymbol{y}|\boldsymbol{\nu}, \boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_n, \sigma_\epsilon^2)$ ▷ Algorithm 1
  Update the fragment for $p(\sigma_\epsilon^2|a_\epsilon)$ ▷ Maestrini and Wand (2020, Algorithm 2)
  Update the fragment for $p(a_\epsilon)$ ▷ Maestrini and Wand (2020, Algorithm 1)
  Update the fragment for $p(\boldsymbol{\nu}|\sigma_\mu^2, \sigma_{\psi_1}^2, \dots, \sigma_{\psi_L}^2)$ ▷ Algorithm 2
  for $i = 1, \dots, n$ **do**
    Update the fragment for $p(\boldsymbol{\zeta}_i)$ ▷ Wand (2017, Section 4.1.1)
  Update the fragment for $p(\sigma_\mu^2|a_\mu)$ ▷ Maestrini and Wand (2020, Algorithm 2)
  Update the fragment for $p(a_\mu)$ ▷ Maestrini and Wand (2020, Algorithm 1)
  for $l = 1, \dots, L$ **do**
    Update the fragment for $p(\sigma_{\psi_l}^2|a_{\psi_l})$ ▷ Maestrini and Wand (2020, Algorithm 2)
    Update the fragment for $p(a_{\psi_l})$ ▷ Maestrini and Wand (2020, Algorithm 1)
Rotate, translate and re-scale $\boldsymbol{\Psi}$ and $\boldsymbol{\Xi}$. ▷ Section 5.2
Outputs: $\widehat{\mu}(\boldsymbol{t}_g)$, $\widehat{\psi}_1(\boldsymbol{t}_g), \dots, \widehat{\psi}_L(\boldsymbol{t}_g)$ and $\widehat{\boldsymbol{\zeta}}_1, \dots, \widehat{\boldsymbol{\zeta}}_n$.

Algorithm 2 Generic VMP algorithm for the MlFPCA model (4.2) with mean field restriction (4.3).

Inputs: All hyperparameters and observed data

Initialize: All factor to stochastic node messages. ▷ Wand (2017, Section 2.5)

Updates:

while $\log \underline{p}(\boldsymbol{y};q)$ has not converged **do**

  Update all stochastic node to factor messages. ▷ Wand (2017, Section 2.5)

  Update the fragment for $p(\boldsymbol{y}|\boldsymbol{\nu},\boldsymbol{\zeta}^{(1)},\boldsymbol{\zeta}^{(2)}_{,}\sigma^2_\epsilon)$ ▷ Algorithm 3

  Update the fragment for $p(\sigma^2_\epsilon|a_\epsilon)$ ▷ Maestrini and Wand (2020, Algorithm 2)

  Update the fragment for $p(a_\epsilon)$ ▷ Maestrini and Wand (2020, Algorithm 1)

  Update the fragment for $p(\boldsymbol{\nu}|\sigma^2_\mu,\boldsymbol{\sigma}^{(1)2}_\psi,\boldsymbol{\sigma}^{(2)2}_\psi)$ ▷ Algorithm 2

  for $i=1,\dots,n$ **do**

    Update the fragment for $p(\boldsymbol{\zeta}_i)$ ▷ (4.8)

  Update the fragment for $p(\sigma^2_\mu|a_\mu)$ ▷ Maestrini and Wand (2020, Algorithm 2)

  Update the fragment for $p(a_\mu)$ ▷ Maestrini and Wand (2020, Algorithm 1)

  for $l=1,\dots,L_1$ **do**

    Update the fragment for $p(\sigma^{(1)2}_{\psi_l}|a^{(1)}_{\psi_l})$ ▷ Maestrini and Wand (2020, Algorithm 2)

    Update the fragment for $p(a^{(1)}_{\psi_l})$ ▷ Maestrini and Wand (2020, Algorithm 1)

  for $l=1,\dots,L_2$ **do**

    Update the fragment for $p(\sigma^{(2)2}_{\psi_l}|a^{(2)}_{\psi_l})$ ▷ Maestrini and Wand (2020, Algorithm 2)

    Update the fragment for $p(a^{(2)}_{\psi_l})$ ▷ Maestrini and Wand (2020, Algorithm 1)

Rotate, translate and re-scale $\boldsymbol{\Psi}^{(1)}$ and $\boldsymbol{\Xi}^{(1)}$. ▷ Section 5.3

Rotate, translate and re-scale $\boldsymbol{\Psi}^{(2)}$ and $\boldsymbol{\Xi}^{(2)}$. ▷ Section 5.3

Outputs: $\widehat{\mu}(\boldsymbol{t}_g)$, $\{\widehat{\psi}^{(1)}_l(\boldsymbol{t}_g)\}_{l=1,\dots,L_1}$, $\{\widehat{\psi}^{(2)}_l(\boldsymbol{t}_g)\}_{l=1,\dots,L_2}$, $\{\widehat{\boldsymbol{\zeta}}^{(1)}_i\}_{i=1,\dots,n}$ and $\{\widehat{\boldsymbol{\zeta}}^{(2)}_{ij}\}_{j=1,\dots,m_i;i=1,\dots,n}$.

References

Maestrini, L. and Wand, M. P. (2020). "The Inverse G-Wishart distribution and variational message passing." *arXiv e-prints*, arXiv:2005.09876. 8, 14, 19, 25, 26

Nolan, T. H., Menictas, M., and Wand, M. P. (2020). "Streamlined computing for variational inference with higher level random effects." *Journal of Machine Learning Research*, 21: 1–62. 16, 17, 18

Nolan, T. H. and Wand, M. P. (2020). "Streamlined Solutions to Multilevel Sparse Matrix Problems." *ANZIAM Journal*, 62: 18–41. 16, 18

Ormerod, J. T. and Wand, M. P. (2010). "Explaining variational approximations." *The American Statistician*, 64: 140–153. 24

Pinheiro, J. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer. 16

Rohde, D. and Wand, M. P. (2016). "Semiparametric mean field variational Bayes: General principles and numerical issues." *Journal of Machine Learning Research*, 17: 1–47. 24

Wand, M. P. (2017). "Fast approximate inference for arbitrarily large semiparametric regression models via message passing (with discussion)." *Journal of the American Statistical Association*, 112: 137–168. 3, 7, 9, 10, 11, 12, 14, 15, 17, 20, 21, 22, 25, 26