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

May 5, 2021

A Proof of Theorem 2.1

We first note that

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

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

$$\phi_1 \equiv \psi_1, \quad \phi_j \equiv \psi_j - \sum_{l=1}^{j-1} \frac{\langle \phi_l, \psi_j \rangle}{||\phi_l||^2}\phi_l, \quad j = 2,\dots,L.$$

Next, set

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

Then $\phi_1^*,\dots,\phi_L^*$ form an orthonormal basis for the span of $\psi_1,\dots,\psi_L$. Therefore, (A.1) can be rewritten as

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

where

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

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

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

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

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

$$\zeta_i^* \equiv \boldsymbol{Q}_\iota^\intercal \iota_i \overset{\text{ind.}}{\sim} \mathrm{N}(0, \Lambda_\iota), \quad i = 1, \dots, n.$$

That is, the elements of ζ_i^* are uncorrelated and $\zeta_1^*, \dots, \zeta_n^*$ are independent.

Next, define the eigenvectors of Σ_ι as $\boldsymbol{q}_1, \dots, \boldsymbol{q}_L$, such that $\boldsymbol{Q} = \begin{bmatrix} q_1 & \cdots & q_L \end{bmatrix}$. Furthermore, define the elements of each of the eigenvectors such that $\boldsymbol{q}_l = (q_{l1}, \dots, q_{lL})^\intercal$, $l = 1, \dots, L$. Then, set

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

The orthonormality of $\psi_1^*, \dots, \psi_L^*$ is easily verified:

$$\langle \psi_l^*, \psi_j^* \rangle = \left\langle \sum_{m=1}^{L} q_{lm} \phi_m^*, \sum_{k=1}^{L} q_{jk} \phi_k^* \right\rangle = \sum_{m=1}^{L} \sum_{k=1}^{L} q_{lm} q_{jk} \langle \phi_m^*, \phi_k^* \rangle = \sum_{k=1}^{L} q_{lk} q_{jk} = \boldsymbol{q}_l^\intercal \boldsymbol{q}_j = \mathbb{I}(l = j),$$

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

Finally, we have

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

The assumptions in the text ensure that this decomposition is unique.

B Proof of Lemma 4.1

To prove (26), we first note that posterior curve estimates from the VMP algorithm satisfy

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

where $\widetilde{\boldsymbol{\zeta}}_i \equiv (\widetilde{\zeta}_{i1}, \dots, \widetilde{\zeta}_{iL})^\intercal$, $i = 1, \dots, n$. Next, define

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

Then, (B.1) implies

$$\boldsymbol{Y} - \mu^*(\boldsymbol{t}_g)\mathbf{1}_N^\intercal = \widetilde{\boldsymbol{\Psi}}\widetilde{\boldsymbol{\Xi}}^\intercal.$$

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

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

It is easy to see that this implies (26).

C Proof of Proposition 4.2

The independence of $\widehat{\boldsymbol{\zeta}}_1,\dots,\widehat{\boldsymbol{\zeta}}_n$ is a consequence of the independence assumption in (8). Let $\boldsymbol{c}$ and $1/\boldsymbol{c}$ retain their definitions from Appendix B. Then, note that

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

Recall that $\boldsymbol{m}_\zeta$ is the mean vector of the columns of $\boldsymbol{D}_\psi\boldsymbol{V}_\psi^\intercal\boldsymbol{\Xi}^\intercal$. Then, it is easy to see that $\sum_{i=1}^n \widehat{\boldsymbol{\zeta}}_i = \mathbf{0}$. Next,

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

which proves the results for the estimated scores.

From Lemma 4.1, we have

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

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 $\widehat{y}_1(\boldsymbol{t}_g),\dots,\widehat{y}_n(\boldsymbol{t}_g)$ is such that

$$\sum_{i=1}^n [\widehat{y}_i(\boldsymbol{t}_g) - \widehat{\mu}(\boldsymbol{t}_g)]\,[\widehat{y}_i(\boldsymbol{t}_g) - \widehat{\mu}(\boldsymbol{t}_g)]^\intercal$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \left(\sum_{l=1}^{L} \widehat{\zeta}_{il} \widehat{\psi}_l(\boldsymbol{t}_g) \right) \left(\sum_{l=1}^{L} \widehat{\zeta}_{il} \widehat{\psi}_l(\boldsymbol{t}_g) \right)^{\mathsf{T}} \\
&= \sum_{i=1}^{n} \left(\widehat{\boldsymbol{\Psi}} \widehat{\boldsymbol{\zeta}}_i \right) \left(\widehat{\boldsymbol{\Psi}} \widehat{\boldsymbol{\zeta}}_i \right)^{\mathsf{T}} \\
&= \widehat{\boldsymbol{\Psi}} \left\{ \sum_{i=1}^{n} \left(\widehat{\boldsymbol{\zeta}}_i \boldsymbol{\zeta}_i^{*\mathsf{T}} \right) \right\} \widehat{\boldsymbol{\Psi}}^{\mathsf{T}} \\
&= (n-1) \widehat{\boldsymbol{\Psi}} \operatorname{diag}(\boldsymbol{c}^2) \widehat{\boldsymbol{\Psi}}^{\mathsf{T}}.
\end{aligned}
$$

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 $\widehat{\psi}_1(\boldsymbol{t}_g), \ldots, \widehat{\psi}_L(\boldsymbol{t}_g)$.

D Derivation of the Functional Principal Component Gaussian Likelihood Fragment

From (6), we have, for $i = 1, \ldots, n$,

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

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

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

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

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

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

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

Next, we consider the updates for standard expectations that occur for each of the random variables and random vectors in (D.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

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

which is a multivariate normal density function with natural parameter vector $\boldsymbol{\eta}_{q^*(\boldsymbol{\nu})}$. From (13), we have

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

Furthermore, the mean vector has the form

$$
\mathbb{E}_q(\boldsymbol{\nu}) \equiv \left\{\mathbb{E}_q(\boldsymbol{\nu}_\mu)^\mathsf{T}, \mathbb{E}_q(\boldsymbol{\nu}_{\psi_1})^\mathsf{T}, \dots, \mathbb{E}_q(\boldsymbol{\nu}_{\psi_L})^\mathsf{T}\right\}^\mathsf{T}, \tag{D.3}
$$

and the covariance matrix has the form

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

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 $\mathbb{C}\text{ov}_q(\boldsymbol{\zeta}_i)$, respectively. The expectations are taken with respect to the normalization of

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

which is a multivariate normal density function with natural parameter vector $\boldsymbol{\eta}_{q^*(\boldsymbol{\zeta}_i)}$. According to (14),

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

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

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

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

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

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

$$C_i\left(\nu_\mu-\sum_{l=1}^{L}\zeta_{il}\nu_{\psi_l}\right)=(\widetilde{\zeta}_i^\mathsf{T}\otimes C_i)\nu. \tag{D.6}$$

Therefore, as a function of ν, (D.1) can be re-written as

$$\begin{aligned}\log p(y_i|\nu,\zeta_i,\sigma_\varepsilon^2)&=-\frac{1}{2\sigma_\varepsilon^2}\left|\left|y_i-(\widetilde{\zeta}_i^\mathsf{T}\otimes C_i)\nu\right|\right|^2+\text{terms not involving }\nu\\&=\begin{bmatrix}\nu\\\text{vec}(\nu\nu^\mathsf{T})\end{bmatrix}^\mathsf{T}\begin{bmatrix}\frac{1}{\sigma_\varepsilon^2}(\widetilde{\zeta}_i^\mathsf{T}\otimes C_i)^\mathsf{T}y_i\\-\frac{1}{2\sigma_\varepsilon^2}\text{vec}\left\{(\widetilde{\zeta}_i\widetilde{\zeta}_i^\mathsf{T})\otimes(C_i^\mathsf{T}C_i)\right\}\end{bmatrix}+\text{terms not involving }\nu.\end{aligned}$$

According to equation (8) of Wand (2017), the message from the factor $p(y|\nu,\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)$ to ν is

$$m_{p(y|\nu,\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)\to\nu}(\nu)\propto\exp\left\{\begin{bmatrix}\nu\\\text{vec}(\nu\nu^\mathsf{T})\end{bmatrix}^\mathsf{T}\eta_{p(y|\nu,\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)\to\nu}\right\},$$

which is proportional to a multivariate normal density function. The update for the message's natural parameter vector, in (15), is dependent upon the mean vector and covariance matrix of $\widetilde{\zeta}_i$, which are

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

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

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

Next, notice that

$$\sum_{l=1}^{L}\zeta_{il}\nu_{\psi_l}=V_\psi\zeta_i \tag{D.9}$$

Then, for each $i=1,\ldots,n$, the log-density function in (D.1) can be represented as a function of ζ_i by

$$\begin{aligned}\log p(y_i|\nu,\zeta_i,\sigma_\varepsilon^2)&=-\frac{1}{2\sigma_\varepsilon^2}\left|\left|y_i-C_i\nu_\mu-C_iV_\psi\zeta_i\right|\right|^2+\text{terms not involving }\zeta_i\\&=\begin{bmatrix}\zeta_i\\\text{vech}(\zeta_i\zeta_i^\mathsf{T})\end{bmatrix}^\mathsf{T}\begin{bmatrix}\frac{1}{\sigma_\varepsilon^2}(V_\psi^\mathsf{T}C_i^\mathsf{T}y_i-h_{\mu\psi,i})\\-\frac{1}{2\sigma_\varepsilon^2}D_L^\mathsf{T}\text{vec}(H_{\psi,i})\end{bmatrix}+\text{terms not involving }\zeta_i,\end{aligned}$$

According to equation (8) of Wand (2017), the message from the factor $p(y|\nu,\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)$ to ζ_i is

$$m_{p(y|\nu,\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)\to\zeta_i}(\zeta_i)\propto\exp\left\{\begin{bmatrix}\zeta_i\\\text{vech}(\zeta_i\zeta_i^\mathsf{T})\end{bmatrix}^\mathsf{T}\eta_{p(y|\nu,\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)\to\zeta_i}\right\},$$

which is proportional to a multivariate normal density function. The message's natural parameter vector update, in (16), 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,

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

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

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

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 (D.3) and (D.4), respectively. Finally, $\mathbb{E}_q(\boldsymbol{H}_{\psi,i})$ is an $L\times L$ matrix, with (l,l') component being

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

The final message to consider is the message from $p(\boldsymbol{y}|\boldsymbol{\nu},\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)$ to σ_ε^2. As a function of σ_ε^2, (D.1) takes the form

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

According to equation (8) of Wand (2017), the message from $p(\boldsymbol{y}|\boldsymbol{\nu},\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)$ to σ_ε^2 is

$$m_{p(\boldsymbol{y}|\boldsymbol{\nu},\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)\to\sigma_\varepsilon^2}(\sigma_\varepsilon^2) \propto \exp\left\{\begin{bmatrix}\log(\sigma_\varepsilon^2)\\ 1/\sigma_\varepsilon^2\end{bmatrix}^\intercal \boldsymbol{\eta}_{p(\boldsymbol{y}|\boldsymbol{\nu},\zeta_1,\ldots,\zeta_n,\sigma_\varepsilon^2)\to\sigma_\varepsilon^2}\right\},$$

which is proportional to an inverse-χ^2 density function. The message's natural parameter vector, in (17), 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

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

where we introduce the matrices

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

and vectors

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

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 (D.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 (D.3). For each $i = 1,\ldots,n$, the expectation of $\boldsymbol{H}_i$, defined in (D.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 (D.14) with expected value

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

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

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

E Derivation of the Functional Principal Component Gaussian Penalization Fragment

From (6), we have, for $l = 1,\ldots,L$,

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

First, we establish the natural parameter vector for each of the optimal posterior density functions. As explained in Appendix D, 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

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

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 (E.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 (13), we have

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

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 (D.3) and (D.4), respectively. For the functional principal components Gaussian penalization fragment, however, we need to note further sub-vectors and sub-matrices. First,

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

and, second,

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

and

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

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 $\boldsymbol{\eta}_{q^*(\sigma^2_\mu)}$. According to Result 6 of Maestrini and Wand (2020),

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

Similar arguments can be used to show that

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

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

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

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

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

which is proportional to a multivariate normal density function. The update for the message's natural parameter vector, in (19), is dependant upon the expectation of $\boldsymbol{\Sigma}_\nu^{-1}$, which is given by

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

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 (E.6) and (E.7), respectively.

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

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

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 σ_μ^2 is

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

which is an inverse-χ^2 density function upon normalization. The message's natural parameter vector update in (20) depends on $\mathbb{E}_q(\boldsymbol{u}_\mu^\mathsf{T}\boldsymbol{u}_\mu)$. Standard statistical results and sub-vector and sub-matrix definitions in (E.3) and (E.4) can be employed to show that

$$\mathbb{E}_q(\boldsymbol{u}_\mu^\mathsf{T}\boldsymbol{u}_\mu) = \mathbb{E}_q(\boldsymbol{u}_\mu)^\mathsf{T}\,\mathbb{E}_q(\boldsymbol{u}_\mu) + \mathrm{tr}\left\{\mathbb{C}\mathrm{ov}_q(\boldsymbol{u}_\mu)\right\}.$$

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

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

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

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

which is an inverse-χ^2 density function upon normalization. The message's natural parameter vector update in (21) depends on $\mathbb{E}_q(\boldsymbol{u}_{\psi_l}^\mathsf{T}\boldsymbol{u}_{\psi_l})$. Standard statistical results and sub-vector and sub-matrix definitions in (E.3) and (E.5) can be employed to show that

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

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

References

Maestrini, L., & Wand, M. P. (2020). The Inverse G-Wishart distribution and variational message passing. *arXiv e-prints*, arXiv:2005.09876.

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.