Technical Report for “Detecting and handling outlying trajectories in irregularly sampled functional datasets”

Daniel Gervini
University of Wisconsin–Milwaukee
Department of Mathematical Sciences
P. O. Box 413, Milwaukee, WI 53201
gervini@uwm.edu

May 11, 2009

Abstract

This Technical Report provides additional material to the paper.

1. Introduction

This Technical Report provides additional material that was omitted in the paper for reasons of space and readability, specifically:

1. a more detailed report of the simulation results;
2. derivation of the maximum likelihood estimating equations;
3. derivation of the EM algorithm steps;
4. proofs of the theorems (including an explicit expression for the matrix $\mathbf{M}$);

2. Simulation Results

This section shows some simulation results that were omitted in the paper:

- Boxplots of $\|\hat{\mu} - \mu\|$ and $\|\hat{\phi}_1 - \phi_1\|$ for all the non-contaminated sampling situations (the three designs and the three sample sizes described in the paper) are shown in Figs. 1 and 2.
- Boxplots of $\|\hat{\mu} - \mu\|$ for contaminated samples, with both $K = 4$ and $K = 8$, are shown in Figs. 3 and 4 (for endogenous and exogenous outliers, respectively).
- Boxplots of $\|\hat{\phi}_1 - \phi_1\|$ for contaminated samples, with both $K = 4$ and $K = 8$, are shown in Figs. 5 and 6 (for endogenous and exogenous outliers, respectively). Figure 7 shows barplots of $\|\hat{\phi}_1 - \phi_1\|$ for endogenously contaminated samples, which make the bimodal nature of the distribution more evident (for better visualization, only the results for $\varepsilon = .10$ and $\varepsilon = .20$ are shown, and only for the Normal and Cauchy estimators).

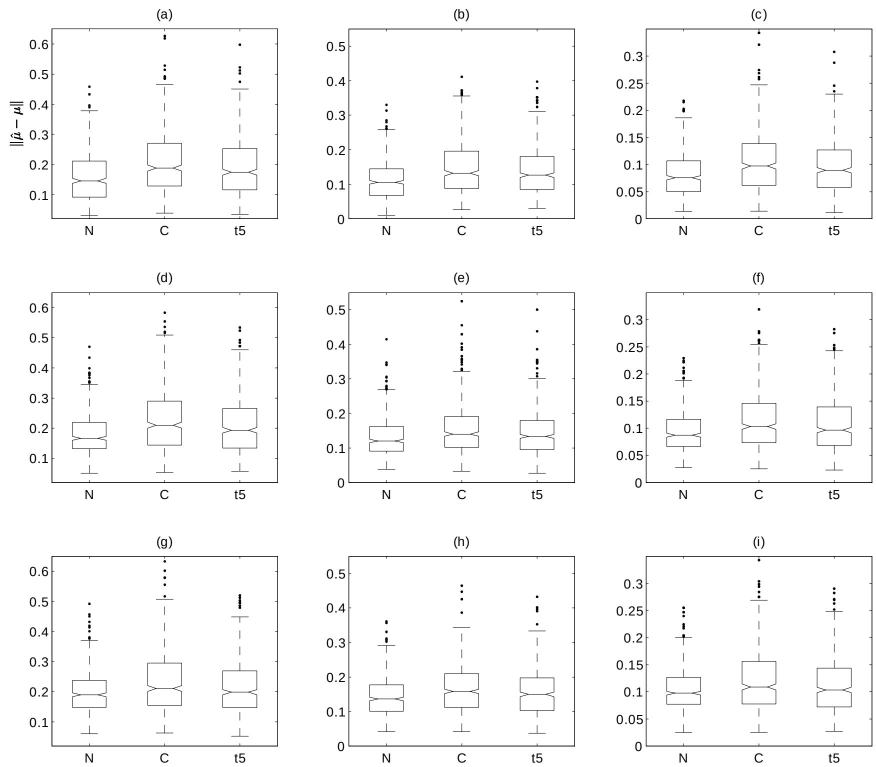

Figure 1: Simulation Results. Box-plots of estimation errors of $\hat{\mu}$ under the normal model, for Normal (N), Cauchy (C) and t_5 (t5) estimators. Designs are fixed uniform (a,b,c), random uniform with fixed m (d,e,f), and random uniform with varying m (g,h,i). Sample sizes are 50 (a,d,g), 100 (b,e,h) and 200 (c,f,i).

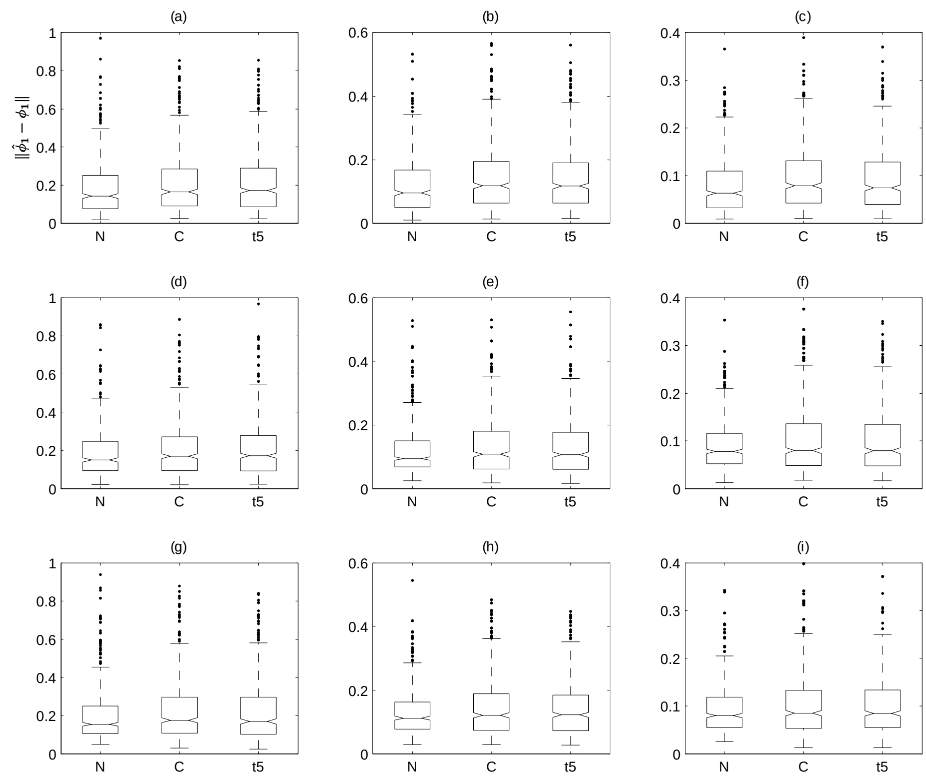

Figure 2: Simulation Results. Box-plots of estimation errors of $\hat{\phi}_1$ under the normal model, for Normal (N), Cauchy (C) and t_5 (t5) estimators. Designs are fixed uniform (a,b,c), random uniform with fixed m (d,e,f), and random uniform with varying m (g,h,i). Sample sizes are 50 (a,d,g), 100 (b,e,h) and 200 (c,f,i).

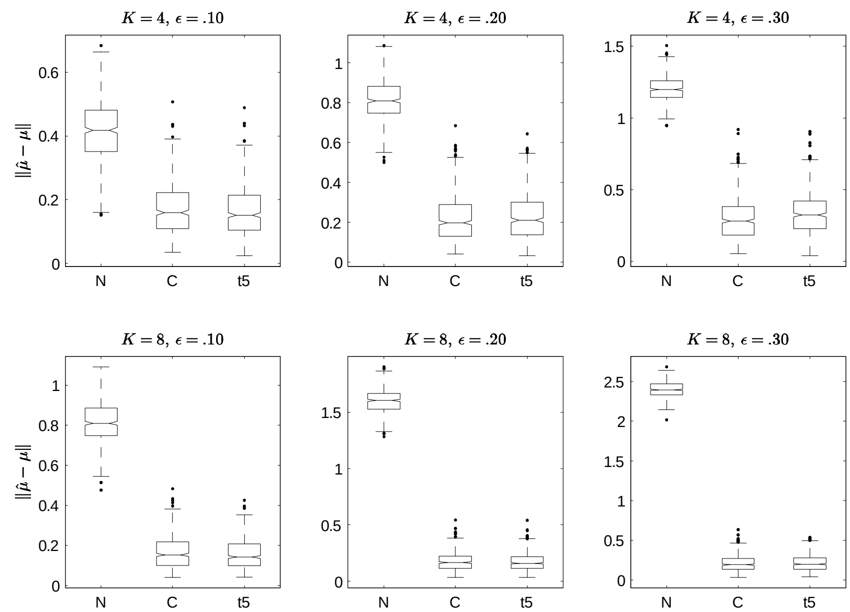

Figure 3: Simulation Results. Box-plots of estimation errors of $\hat{\mu}$ under the contaminated-normal model (with endogenous outliers), for Normal (N), Cauchy (C) and t_5 (t5) estimators.

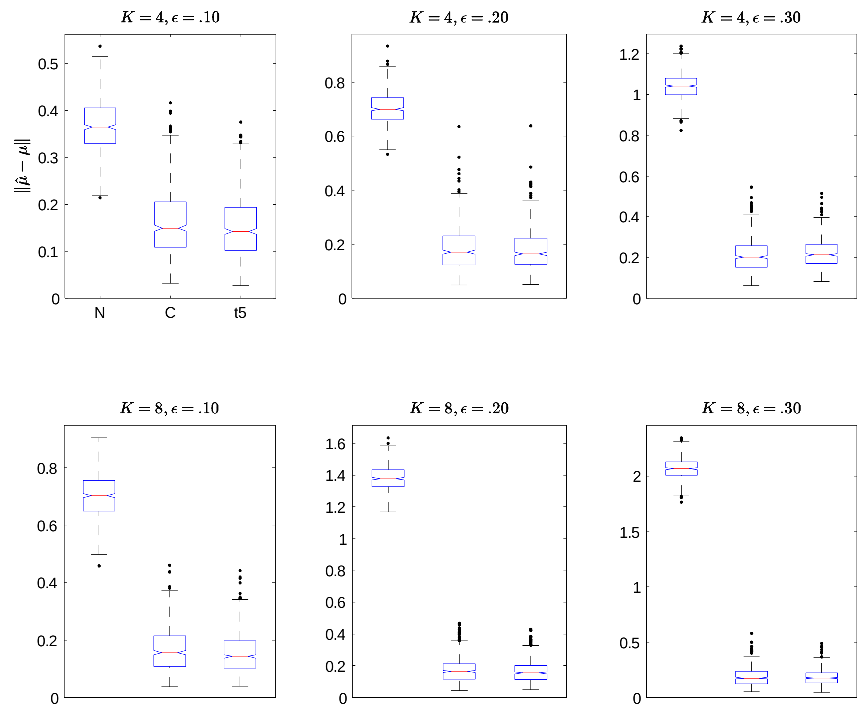

Figure 4: Simulation Results. Box-plots of estimation errors of $\hat{\mu}$ under the contaminated-normal model (with exogenous outliers), for Normal (N), Cauchy (C) and t_5 (t5) estimators.

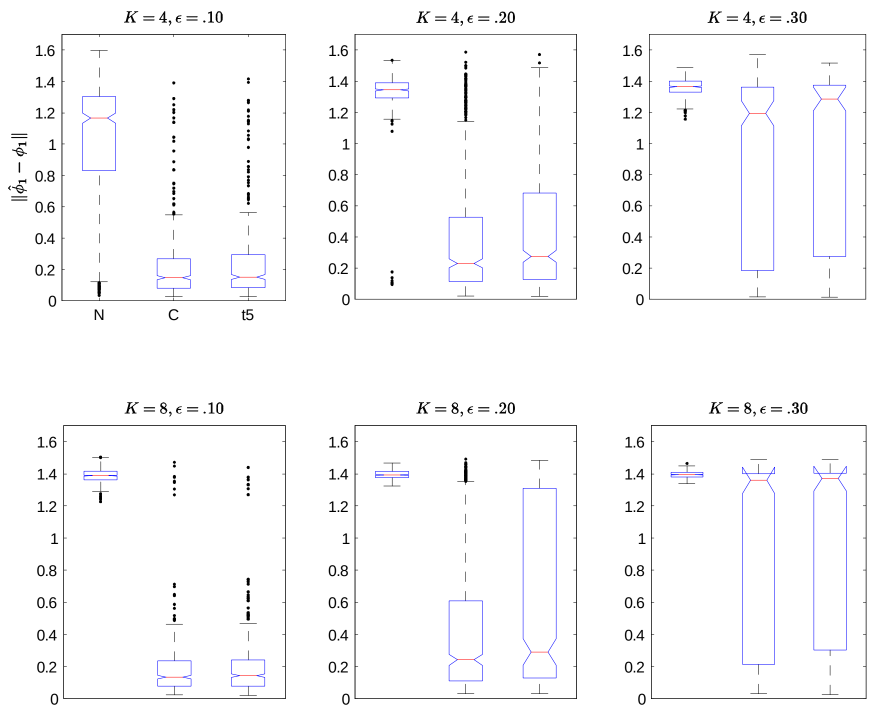

Figure 5: Simulation Results. Box-plots of estimation errors of $\hat{\phi}_1$ under the contaminated-normal model (with endogenous outliers), for Normal (N), Cauchy (C) and t_5 (t5) estimators.

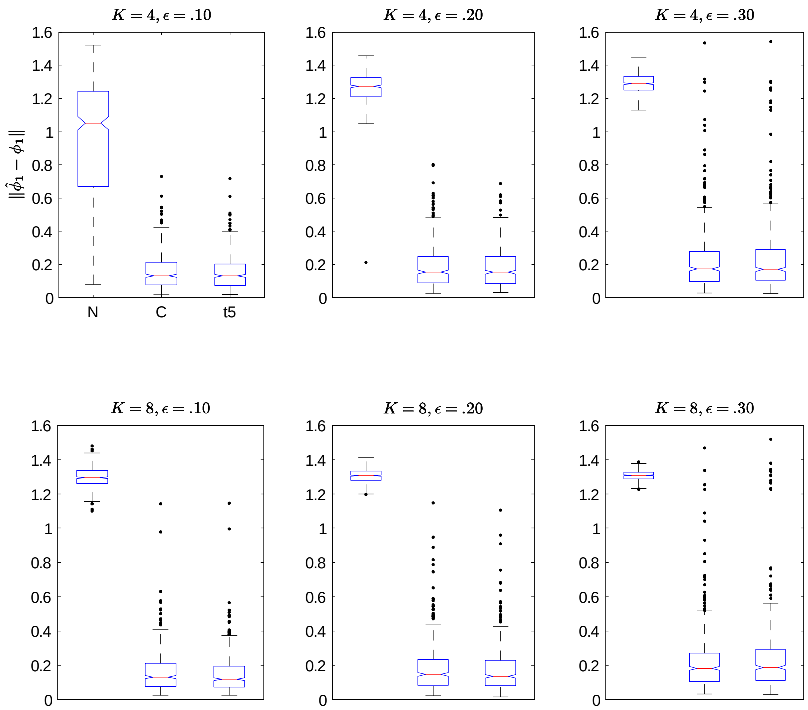

Figure 6: Simulation Results. Box-plots of estimation errors of $\hat{\phi}_1$ under the contaminated-normal model (with exogenous outliers), for Normal (N), Cauchy (C) and t_5 (t5) estimators.

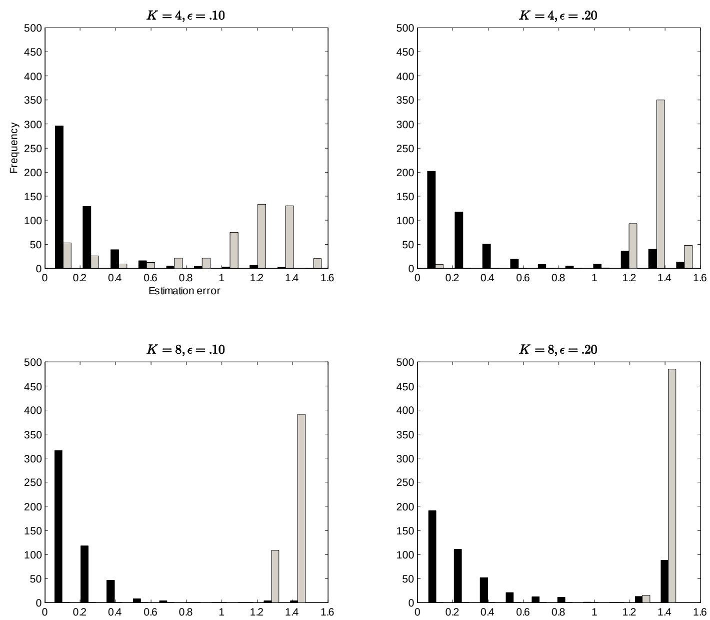

Figure 7: Simulation Results. Bar plots of estimation errors of $\hat{\phi}_1$ under the contaminated Normal model (with endogenous outliers), for Normal (light grey) and Cauchy (black) estimators.

3. Maximum likelihood estimating equations

The density function of a $t_\nu(\boldsymbol{\mu}, \Sigma)$ distribution in $\mathbb{R}^m$ is

$$f(\mathbf{x}) = \frac{\Gamma\{(\nu+m)/2\}}{\Gamma(\nu/2)}(\nu\pi)^{-m/2}\,|\Sigma|^{-1/2}\left(1+\frac{s}{\nu}\right)^{-(\nu+m)/2},$$

with $s = (\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$. Then

$$\log f(\mathbf{x}) \propto -\frac{1}{2}\log|\Sigma| - \left(\frac{\nu+m}{2}\right)\log\left(1+\frac{s}{\nu}\right).$$

To find the derivatives of $\log f(\mathbf{x})$ with respect to the parameters we use differentials, as in Magnus and Neudecker (1999).

Differentiating with respect to $\boldsymbol{\mu}$ we obtain

$$\mathrm{d}\log f(\mathbf{x}) = -\left(\frac{\nu+m}{2}\right)\left(1+\frac{s}{\nu}\right)^{-1}\frac{1}{\nu}\mathrm{d}s$$

and

$$\mathrm{d}s = -2(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}\mathrm{d}\boldsymbol{\mu}.$$

Therefore

$$\mathrm{d}\log f(\mathbf{x}) = \left(\frac{\nu+m}{\nu+s}\right)(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}\mathrm{d}\boldsymbol{\mu}.$$

If $\boldsymbol{\mu} = B\boldsymbol{\theta}$, $\mathrm{d}\boldsymbol{\mu} = B\mathrm{d}\boldsymbol{\theta}$ and then

$$\frac{\partial\log f(\mathbf{x})}{\partial\boldsymbol{\theta}^T} = \left(\frac{\nu+m}{\nu+s}\right)(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B.$$

Differentiating with respect to Σ, and using that

$$\begin{aligned}\mathrm{d}\log|\Sigma| &= \mathrm{tr}\left(\Sigma^{-1}\mathrm{d}\Sigma\right),\\ \mathrm{d}(\Sigma^{-1}) &= -\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1},\end{aligned}$$

we have

$$\mathrm{d}\log f(\mathbf{x}) = -\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\mathrm{d}\Sigma\right) - \left(\frac{\nu+m}{2}\right)\left(1+\frac{s}{\nu}\right)^{-1}\frac{1}{\nu}\mathrm{d}s$$

and

$$\mathrm{d}s = -(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}),$$

so

$$\mathrm{d}\log f(\mathbf{x}) = -\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\mathrm{d}\Sigma\right) + \frac{1}{2}\left(\frac{\nu+m}{\nu+s}\right)(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}).$$

For $\Sigma = \sigma^2 I_m + BH\Lambda H^T B^T$ we have

$$\begin{aligned}
\mathrm{d}\Sigma &= I_m \mathrm{d}\sigma^2, \\
\mathrm{d}\Sigma &= B(\mathrm{d}H)\Lambda H^T B^T + BH\Lambda(\mathrm{d}H)^T B^T, \\
\mathrm{d}\Sigma &= BH(\mathrm{d}\Lambda)H^T B^T.
\end{aligned}$$

Then
$$\frac{\partial \log f(\mathbf{x})}{\partial \sigma^2} = -\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\right) + \frac{1}{2}\left(\frac{\nu+m}{\nu+s}\right)(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-2}(\mathbf{x}-\boldsymbol{\mu}).$$

Note that $\mathrm{d}\boldsymbol{\eta}_k$ is the kth column of $\mathrm{d}H$, so $\mathrm{d}H = \sum_{k=1}^d (\mathrm{d}\boldsymbol{\eta}_k)\mathbf{e}_k^T$, where $\mathbf{e}_k$ is the kth canonical vector in $\mathbb{R}^d$. Then

$$\begin{aligned}
\mathrm{d}\Sigma &= \sum_{k=1}^d \{B(\mathrm{d}\boldsymbol{\eta}_k)\mathbf{e}_k^T\Lambda H^T B^T + BH\Lambda\mathbf{e}_k(\mathrm{d}\boldsymbol{\eta}_k)^T B^T\} \\
&= \sum_{k=1}^d \{B(\mathrm{d}\boldsymbol{\eta}_k)\lambda_k\boldsymbol{\eta}_k^T B^T + B\boldsymbol{\eta}_k\lambda_k(\mathrm{d}\boldsymbol{\eta}_k)^T B^T\},
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathrm{d}\log f(\mathbf{x}) &= \sum_{k=1}^d \{-\lambda_k\boldsymbol{\eta}_k^T B^T\Sigma^{-1}B(\mathrm{d}\boldsymbol{\eta}_k) \\
&\quad + \left(\frac{\nu+m}{\nu+s}\right)(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B(\mathrm{d}\boldsymbol{\eta}_k)\lambda_k\boldsymbol{\eta}_k^T B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}.
\end{aligned}$$

Then

$$\frac{\partial \log f(\mathbf{x})}{\partial \boldsymbol{\eta}_k^T} = -\lambda_k\boldsymbol{\eta}_k^T B^T\Sigma^{-1}B + \left(\frac{\nu+m}{\nu+s}\right)\lambda_k\boldsymbol{\eta}_k^T B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B,$$

or more succinctly
$$\frac{\partial \log f(\mathbf{x})}{\partial \boldsymbol{\eta}_k^T} = \lambda_k\boldsymbol{\eta}_k^T S,$$

with

$$S = -B^T\Sigma^{-1}B + \left(\frac{\nu+m}{\nu+s}\right) B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B.$$

Similarly, since $\mathrm{d}\Lambda = \sum_{k=1}^{d}(\mathrm{d}\lambda_k)\mathbf{e}_k\mathbf{e}_k^T$, we have

$$\mathrm{d}\Sigma = \sum_{k=1}^{d}(\mathrm{d}\lambda_k)B\boldsymbol{\eta}_k\boldsymbol{\eta}_k^T B^T$$

and then

$$\begin{aligned}\mathrm{d}\log f(\mathbf{x}) &= \sum_{k=1}^{d}\{-\frac{1}{2}\boldsymbol{\eta}_k^T B^T\Sigma^{-1}B\boldsymbol{\eta}_k \mathrm{d}\lambda_k \\ &\quad +\frac{1}{2}\left(\frac{\nu+m}{\nu+s}\right)(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathrm{d}\lambda_k)B\boldsymbol{\eta}_k\boldsymbol{\eta}_k^T B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\},\end{aligned}$$

which implies that

$$\frac{\partial \log f(\mathbf{x})}{\partial\lambda_k} = \frac{1}{2}\boldsymbol{\eta}_k^T S\boldsymbol{\eta}_k.$$

To compute the derivatives of the Lagrangian function, note that

$$\frac{1}{2}\sum_{j=1}^{d} w_{jj}(1-\boldsymbol{\eta}_j^T J\boldsymbol{\eta}_j) - \sum_{j=2}^{d}\sum_{k=1}^{j-1} w_{jk}\boldsymbol{\eta}_j^T J\boldsymbol{\eta}_k = \frac{1}{2}\mathrm{tr}\{W(I_d - H^T JH)\},$$

with W symmetric and $W_{ij} = w_{ij}$. Then the differential with respect to H is

$$\begin{aligned}-\frac{1}{2}\mathrm{tr}\{W(\mathrm{d}H)^T JH + WH^T J(\mathrm{d}H)\} &= -\mathrm{tr}\{WH^T J(\mathrm{d}H)\} \\ &= -\sum_{k=1}^{d}\mathbf{e}_k^T WH^T J(\mathrm{d}\boldsymbol{\eta}_k).\end{aligned}$$

This implies that

$$\frac{\partial L(\boldsymbol{\xi})}{\partial\boldsymbol{\eta}_k^T} = \lambda_k\boldsymbol{\eta}_k^T S_n - \mathbf{e}_k^T WH^T J.$$

Since this is equal to zero at $\hat{\boldsymbol{\xi}}$ and $H^T JH = I_d$, the multiplier matrix W satisfies the equation

$$\lambda_k\boldsymbol{\eta}_k^T S_n H - \mathbf{e}_k^T W = \mathbf{0},$$

from which we derive the estimating equation

$$\lambda_k \boldsymbol{\eta}_k^T S_n - \lambda_k \boldsymbol{\eta}_k^T S_n H H^T J = \mathbf{0}.$$

4. Derivation of the EM algorithm steps

Under the central model,

$$\begin{aligned}\mathbf{x}_i &= B_i\boldsymbol{\theta} + B_i\Xi\mathbf{z}_i + \sigma\boldsymbol{\varepsilon}_i \\ &= B_i\boldsymbol{\theta} + (\sigma I_{m_i}, B_i\Xi)\mathbf{v}_i,\end{aligned}$$

where $\mathbf{v}_i = (\boldsymbol{\varepsilon}_i^T, \mathbf{z}_i^T)^T$ is assumed to have a $t_\nu(\mathbf{0}, I_{m_i+d})$ distribution. Then $\mathbf{v}_i = \tilde{\mathbf{v}}_i/\sqrt{u_i}$, where $\tilde{\mathbf{v}}_i \sim N(\mathbf{0}, I_{m_i+d})$, $u_i \sim \nu^{-1}\chi^2_\nu$, and $\tilde{\mathbf{v}}_i$ and u_i are independent. Note that $u_i \sim \Gamma(\nu/2, 2/\nu)$, so

$$f(u_i) \propto \; u_i^{\nu/2-1}\exp(-\frac{u_i\nu}{2}).$$

If we decompose $\tilde{\mathbf{v}}_i$ as $(\tilde{\boldsymbol{\varepsilon}}_i^T, \tilde{\mathbf{z}}_i^T)^T$ we have that $\tilde{\mathbf{z}}_i \sim N(\mathbf{0}, I_d)$, $\tilde{\boldsymbol{\varepsilon}}_i \sim N(\mathbf{0}, I_{m_i})$, and $\tilde{\mathbf{z}}_i$ and $\tilde{\boldsymbol{\varepsilon}}_i$ are independent. Then

$$\mathbf{x}_i|(\tilde{\mathbf{z}}_i, u_i) \sim N(B_i\boldsymbol{\theta} + B_i\Xi\tilde{\mathbf{z}}_i/\sqrt{u_i}, \sigma^2 I_{m_i}/u_i).$$

Treating $\{(\tilde{\mathbf{z}}_i, u_i)\}$ as "missing data," the joint density function factors as $f(\mathbf{x}_i, \tilde{\mathbf{z}}_i, u_i) = f(\mathbf{x}_i|\tilde{\mathbf{z}}_i, u_i)f(\tilde{\mathbf{z}}_i)f(u_i)$ and the complete-data likelihood is

$$\begin{aligned}l^*(\boldsymbol{\xi}) &= \sum_{i=1}^n \log f(\mathbf{x}_i, \tilde{\mathbf{z}}_i, u_i) \\ &\propto \sum_{i=1}^n \{-\frac{m_i}{2}\log\left(\sigma^2\right) - \frac{u_i}{2\sigma^2}\|\mathbf{x}_i - B_i\boldsymbol{\theta} - \frac{1}{\sqrt{u_i}}B_i\Xi\tilde{\mathbf{z}}_i\|^2 + h(\tilde{\mathbf{z}}_i, u_i)\} \\ &= \sum_{i=1}^n \{-\frac{m_i}{2}\log\left(\sigma^2\right) - \frac{u_i}{2\sigma^2}\left\|\mathbf{x}_i - B_i\boldsymbol{\theta}\right\|^2 - \frac{1}{2\sigma^2}\left\|B_i\Xi\tilde{\mathbf{z}}_i\right\|^2 \\ &\quad + \frac{\sqrt{u_i}}{\sigma^2}(\mathbf{x}_i - B_i\boldsymbol{\theta})^T B_i\Xi\tilde{\mathbf{z}}_i + h(\tilde{\mathbf{z}}_i, u_i)\},\end{aligned}$$

where $h(\tilde{\mathbf{z}}_i, u_i)$ is a function that does not depend on $\boldsymbol{\xi}$. The expectation step of the EM algorithm defines $Q(\boldsymbol{\xi}) = \mathrm{E}^{\mathrm{old}}\{l^*(\boldsymbol{\xi})|\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, where $\mathrm{E}^{\mathrm{old}}$ denotes expectation with respect to the current parameters $\hat{\boldsymbol{\xi}}^{\rm old}$. The maximization step of the EM algorithm defines $\hat{\boldsymbol{\xi}}^{\rm new}$ as the maximizer of $Q(\boldsymbol{\xi})$.

Explicitly,

$$
\begin{aligned}
Q(\boldsymbol{\xi}) \;\propto\; & \sum_{i=1}^{n}[-\frac{m_i}{2}\log\left(\sigma^2\right) - \frac{\mathrm{E}^{\rm old}(u_i|\mathbf{x}_i)}{2\sigma^2}\|\mathbf{x}_i - B_i\boldsymbol{\theta}\|^2 \\
& -\frac{1}{2\sigma^2}\mathrm{tr}\{B_i\Xi\mathrm{E}^{\rm old}(\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T|\mathbf{x}_i)\Xi^T B_i^T\} \\
& +\frac{1}{\sigma^2}(\mathbf{x}_i - B_i\boldsymbol{\theta})^T B_i\Xi\mathrm{E}^{\rm old}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i) + \mathrm{E}^{\rm old}\{h(\tilde{\mathbf{z}}_i, u_i)|\mathbf{x}_i\}].
\end{aligned}
$$

Since $\mathrm{tr}(ABCD) = \mathrm{vec}(B^T)^T(A^T \otimes C)\mathrm{vec}(D)$ (see e.g. Magnus and Neudecker 1999), we have

$$
\begin{aligned}
\mathrm{tr}\{B_i\Xi\mathrm{E}^{\rm old}(\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T|\mathbf{x}_i)\Xi^T B_i^T\} &= \mathrm{tr}\{\mathrm{E}^{\rm old}(\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T|\mathbf{x}_i)\Xi^T B_i^T B_i\Xi\} \\
&= \mathrm{vec}(\Xi)^T\{\mathrm{E}^{\rm old}(\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T|\mathbf{x}_i)\otimes B_i^T B_i\}\mathrm{vec}(\Xi), \\
(\mathbf{x}_i - B_i\boldsymbol{\theta})^T B_i\Xi\mathrm{E}^{\rm old}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i) &= \mathrm{tr}\{\mathrm{E}^{\rm old}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)(\mathbf{x}_i - B_i\boldsymbol{\theta})^T B_i\Xi\} \\
&= (\mathbf{x}_i - B_i\boldsymbol{\theta})^T\{\mathrm{E}^{\rm old}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)^T\otimes B_i\}\mathrm{vec}(\Xi).
\end{aligned}
$$

Then

$$
\frac{\partial Q}{\partial\boldsymbol{\theta}^T} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\{\mathrm{E}^{\rm old}(u_i|\mathbf{x}_i)(\mathbf{x}_i - B_i\boldsymbol{\theta})^T B_i - \mathrm{E}^{\rm old}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)^T\Xi^T B_i^T B_i\}, \tag{1}
$$

$$
\begin{aligned}
\frac{\partial Q}{\partial\mathrm{vec}(\Xi)^T} \;=\; & \frac{1}{\sigma^2}\sum_{i=1}^{n}[-\mathrm{vec}(\Xi)^T\{\mathrm{E}^{\rm old}(\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T|\mathbf{x}_i)\otimes B_i^T B_i\} \\
& +(\mathbf{x}_i - B_i\boldsymbol{\theta})^T\{\mathrm{E}^{\rm old}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)^T\otimes B_i\}],
\end{aligned} \tag{2}
$$

and

$$
\begin{aligned}
\frac{\partial Q}{\partial\sigma^2} \;=\; & \sum_{i=1}^{n}[-\frac{m_i}{2\sigma^2} + \frac{\mathrm{E}^{\rm old}(u_i|\mathbf{x}_i)}{2(\sigma^2)^2}\|\mathbf{x}_i - B_i\boldsymbol{\theta}\|^2 \\
& +\frac{1}{2(\sigma^2)^2}\mathrm{tr}\{B_i\Xi\mathrm{E}^{\rm old}(\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T|\mathbf{x}_i)\Xi^T B_i^T\} \\
& -\frac{1}{(\sigma^2)^2}(\mathbf{x}_i - B_i\boldsymbol{\theta})^T B_i\Xi\mathrm{E}^{\rm old}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)].
\end{aligned} \tag{3}
$$

For simplicity, we solve (1) and (2) separately. That is, we replace Ξ by $\hat{\Xi}^{\text{old}}$ on the right-hand side of (1) and $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}^{\text{old}}$ on the right-hand side of (2). Then we obtain

$$\begin{aligned}\hat{\boldsymbol{\theta}}^{\text{new}} &= \left\{\sum_{i=1}^{n} \mathrm{E}^{\text{old}}(u_i|\mathbf{x}_i) B_i^T B_i\right\}^{-1} \\ &\times \sum_{i=1}^{n} \{\mathrm{E}^{\text{old}}(u_i|\mathbf{x}_i) B_i^T \mathbf{x}_i - B_i^T B_i \hat{\Xi}^{\text{old}} \mathrm{E}^{\text{old}}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)\},\end{aligned}$$

$$\begin{aligned}\operatorname{vec}(\hat{\Xi}^{\text{new}}) &= \left[\sum_{i=1}^{n} \{\mathrm{E}^{\text{old}}(\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^T|\mathbf{x}_i) \otimes B_i^T B_i\}\right]^{-1} \\ &\times \sum_{i=1}^{n} \{\mathrm{E}^{\text{old}}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i) \otimes B_i^T\}(\mathbf{x}_i - B_i \hat{\boldsymbol{\theta}}^{\text{old}}),\end{aligned}$$

and

$$\begin{aligned}(\hat{\sigma}^2)^{\text{new}} &= \frac{1}{\sum_{i=1}^{n} m_i} \sum_{i=1}^{n} [\mathrm{E}^{\text{old}}(u_i|\mathbf{x}_i) \|\mathbf{x}_i - B_i \hat{\boldsymbol{\theta}}^{\text{new}}\|^2 \\ &+ \operatorname{tr}\{B_i \hat{\Xi}^{\text{new}} \mathrm{E}^{\text{old}}(\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^T|\mathbf{x}_i)(\hat{\Xi}^{\text{new}})^T B_i^T\} \\ &- 2(\mathbf{x}_i - B_i \hat{\boldsymbol{\theta}}^{\text{new}})^T B_i \hat{\Xi}^{\text{new}} \mathrm{E}^{\text{old}}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)].\end{aligned}$$

Note that

$$Q(\hat{\boldsymbol{\xi}}^{\text{new}}) \propto -(\sum_{i=1}^{n} \frac{m_i}{2})[\log\{(\hat{\sigma}^2)^{\text{new}}\} + 1],$$

so $\hat{\boldsymbol{\xi}}$ is the minimizer of $\hat{\sigma}^2 = \sigma^2(\boldsymbol{\xi})$.

Now we derive explicit expressions for $\mathrm{E}(u_i|\mathbf{x}_i)$, $\mathrm{E}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)$ and $\mathrm{E}(\tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^T|\mathbf{x}_i)$. Since $\mathbf{x}_i|u_i \sim N(B_i\boldsymbol{\theta}, \Sigma_i/u_i)$ and $u_i \sim \Gamma(\nu/2, 2/\nu)$, we have

$$f(\mathbf{x}_i, u_i) \propto |\Sigma_i|^{-1/2}\, u_i^{-m_i/2} \exp(-\frac{u_i s_i}{2})\; u_i^{\nu/2-1} \exp(-\frac{u_i \nu}{2}),$$

which implies that $u_i|\mathbf{x}_i \sim \Gamma\{(\nu + m_i)/2, 2/(\nu + s_i)\}$ and then

$$\mathrm{E}(u_i|\mathbf{x}_i) = \frac{\nu + m_i}{\nu + s_i}.$$

To find $\mathrm{E}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i)$ and $\mathrm{E}(\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T|\mathbf{x}_i)$ we need the distribution of $\tilde{\mathbf{z}}_i|(u_i, \mathbf{x}_i)$. Since

$$\begin{bmatrix} \mathbf{x}_i \\ \tilde{\mathbf{z}}_i \end{bmatrix} = \begin{bmatrix} B_i\boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \sigma I_{m_i}/\sqrt{u_i} & B_i\Xi/\sqrt{u_i} \\ 0 & I_d \end{bmatrix} \tilde{\mathbf{v}}_i$$

and $\tilde{\mathbf{v}}_i \sim N(\mathbf{0}, I_{d+m_i})$,

$$\begin{bmatrix} \mathbf{x}_i \\ \tilde{\mathbf{z}}_i \end{bmatrix} \sim N\left(\begin{bmatrix} B_i\boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma^2 I_{m_i}/u_i & B_i\Xi/\sqrt{u_i} \\ \Xi^T B_i^T/\sqrt{u_i} & I_d \end{bmatrix} \right).$$

Well-known properties of the multivariate normal distribution (see e.g. Bilodeau and Brenner 1999) imply that

$$\tilde{\mathbf{z}}_i|(u_i, \mathbf{x}_i) \sim N(\sqrt{u_i}\Xi^T B_i^T \Sigma_i^{-1}(\mathbf{x}_i - B_i\boldsymbol{\theta}), I_d - \Xi^T B_i^T \Sigma_i^{-1} B_i \Xi).$$

Written out more explicitly,

$$I_d - \Xi^T B_i^T \Sigma_i^{-1} B_i \Xi = I_d - \frac{1}{\sqrt{u_i}}\Xi^T B_i^T (\frac{1}{u_i}\Sigma_i)^{-1} \frac{1}{\sqrt{u_i}} B_i \Xi$$

and

$$\frac{1}{u_i}\Sigma_i = \frac{\sigma^2}{u_i} I_{m_i} + \frac{1}{\sqrt{u_i}} B_i \Xi I_d^{-1} \frac{1}{\sqrt{u_i}} \Xi^T B_i^T.$$

Note that $I_d - \Xi^T B_i^T \Sigma_i^{-1} B_i \Xi = V_i^{-1}$, where $V_i = I_d + \Xi^T B_i^T B_i \Xi/\sigma^2$, which follows from the identity

$$(A + UV)^{-1} = A^{-1} - A^{-1}U(I + VA^{-1}U)^{-1}VA^{-1}.$$

Since $f(u_i, \tilde{\mathbf{z}}_i|\mathbf{x}_i) = f(\tilde{\mathbf{z}}_i|u_i, \mathbf{x}_i) f(u_i|\mathbf{x}_i)$,

$$\mathrm{E}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i) = \mathrm{E}(u_i|\mathbf{x}_i)\Xi^T B_i^T \Sigma_i^{-1}(\mathbf{x}_i - B_i\boldsymbol{\theta})$$

and also

$$\begin{aligned} \mathrm{E}(\mathbf{z}_i|\mathbf{x}_i) &= \mathrm{E}(u_i^{-1/2}\tilde{\mathbf{z}}_i|\mathbf{x}_i) \\ &= \Xi^T B_i^T \Sigma_i^{-1}(\mathbf{x}_i - B_i\boldsymbol{\theta}), \end{aligned}$$

so $\hat{\mathbf{z}}_i$ is as claimed. Then

$$\mathrm{E}(\sqrt{u_i}\tilde{\mathbf{z}}_i|\mathbf{x}_i) = \mathrm{E}(u_i|\mathbf{x}_i)\hat{\mathbf{z}}_i$$

and

$$\mathrm{E}(\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T|\mathbf{x}_i) = V_i^{-1} + \mathrm{E}(u_i|\mathbf{x}_i)\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i^T.$$

5. Asymptotic properties

If F_n denotes the empirical distribution of $\mathbf{w}_i = (\mathbf{t}_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, the maximum likelihood estimating equations can be succinctly expressed as $\mathrm{E}_{F_n}\{\boldsymbol{\psi}(\mathbf{w}, \hat{\boldsymbol{\xi}})\} = \mathbf{0}$, where $\boldsymbol{\psi} : \mathbb{R}^{2m} \times \mathbb{R}^{(p+1)(d+1)} \to \mathbb{R}^{(p+1)(d+1)}$ can be broken down as follows:

$$\boldsymbol{\psi}_1(\mathbf{w}, \boldsymbol{\xi}) = \left\{\frac{\nu + m}{\nu + s(\mathbf{w}, \boldsymbol{\xi})}\right\} B^T(\mathbf{t})\Sigma^{-1}(\mathbf{t}, \boldsymbol{\xi})\{\mathbf{x} - B(\mathbf{t})\boldsymbol{\theta}\},$$

$$\boldsymbol{\psi}_2(\mathbf{w}, \boldsymbol{\xi}) = \mathrm{vec}\{(I_p - JHH^T)S(\mathbf{w}, \boldsymbol{\xi})H\},$$

$$\boldsymbol{\psi}_3(\mathbf{w}, \boldsymbol{\xi}) = [\boldsymbol{\eta}_1^T S(\mathbf{w}, \boldsymbol{\xi})\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d^T S(\mathbf{w}, \boldsymbol{\xi})\boldsymbol{\eta}_d]^T,$$

and

$$\begin{aligned}\psi_4(\mathbf{w}, \boldsymbol{\xi}) &= -\frac{1}{2}\mathrm{tr}\{\Sigma^{-1}(\mathbf{t}, \boldsymbol{\xi})\} \\ &\quad + \frac{1}{2}\left\{\frac{\nu + m}{\nu + s(\mathbf{w}, \boldsymbol{\xi})}\right\}\{\mathbf{x} - B(\mathbf{t})\boldsymbol{\theta}\}^T\Sigma^{-2}(\mathbf{t}, \boldsymbol{\xi})\{\mathbf{x} - B(\mathbf{t})\boldsymbol{\theta}\},\end{aligned}$$

where

$$\begin{aligned}S(\mathbf{w}, \boldsymbol{\xi}) &= -B^T(\mathbf{t})\Sigma^{-1}(\mathbf{t}, \boldsymbol{\xi})B(\mathbf{t}) \\ &\quad + \left\{\frac{\nu + m}{\nu + s(\mathbf{w}, \boldsymbol{\xi})}\right\} B^T(\mathbf{t})\Sigma^{-1}(\mathbf{t}, \boldsymbol{\xi})\{\mathbf{x} - B(\mathbf{t})\boldsymbol{\theta}\} \\ &\quad \times \{\mathbf{x} - B(\mathbf{t})\boldsymbol{\theta}\}^T\Sigma^{-1}(\mathbf{t}, \boldsymbol{\xi})B(\mathbf{t}),\end{aligned}$$

$$\Sigma(\mathbf{t}, \boldsymbol{\xi}) = B(\mathbf{t})H\Lambda H^T B^T(\mathbf{t}) + \sigma^2 I_m,$$

and

$$s(\mathbf{w}, \boldsymbol{\xi}) = \{\mathbf{x} - B(\mathbf{t})\boldsymbol{\theta}\}^T\Sigma^{-1}(\mathbf{t}, \boldsymbol{\xi})\{\mathbf{x} - B(\mathbf{t})\boldsymbol{\theta}\}.$$

Let us denote the true model parameters by a null subindex, and let $\boldsymbol{\xi}^* = \boldsymbol{\xi}(F)$. Then $\Sigma(\mathbf{t}, \boldsymbol{\xi}^*) = \beta_0 \Sigma_0(\mathbf{t})$, $s(\mathbf{w}, \boldsymbol{\xi}^*) = s_0(\mathbf{w})/\beta_0$ and $S(\mathbf{w}, \boldsymbol{\xi}^*) = S_0(\mathbf{w})/\beta_0$, where

$$\begin{aligned} S_0(\mathbf{w}) &= -B^T(\mathbf{t})\Sigma_0^{-1}(\mathbf{t})B(\mathbf{t}) \\ &\quad + \left\{\frac{\nu + m}{\beta_0 \nu + s_0(\mathbf{w})}\right\} B^T(\mathbf{t})\Sigma_0^{-1}(\mathbf{t})\{\mathbf{x} - \mu_0(\mathbf{t})\} \\ &\quad \times \{\mathbf{x} - \mu_0(\mathbf{t})\}^T \Sigma_0^{-1}(\mathbf{t}) B(\mathbf{t}). \end{aligned}$$

5.1. Proof of Theorem 1

In this proof we will use certain properties of spherical distributions that can be found, for instance, in Bilodeau and Brenner (1999). Under the model we have $\mathbf{x} = \mu_0(\mathbf{t}) + A(\mathbf{t})\mathbf{v}$, where $A(\mathbf{t}) = [B(\mathbf{t})H_0\Lambda_0^{1/2}, \sigma_0 I_m]$ and $\mathbf{v} = (\mathbf{z}^T, \boldsymbol{\varepsilon}^T)^T$. Note that $\Sigma_0(\mathbf{t}) = A(\mathbf{t})A^T(\mathbf{t})$ and $\operatorname{rank}\{A(\mathbf{t})\} = m$ for all $\mathbf{t}$. Let $A(\mathbf{t})^T = Q(\mathbf{t})R(\mathbf{t})$ be the QR-decomposition of $A(\mathbf{t})^T$, with $Q(\mathbf{t}) \in \mathbb{R}^{(d+m)\times m}$ orthogonal and $R(\mathbf{t}) \in \mathbb{R}^{m\times m}$ invertible (hence $\Sigma_0(\mathbf{t}) = R^T(\mathbf{t})R(\mathbf{t})$.) Since $\mathbf{v}$ has a spherical distribution in $\mathbb{R}^{d+m}$, $\mathbf{y} = Q(\mathbf{t})^T\mathbf{v}$ has a spherical distribution in $\mathbb{R}^m$ (which does not depend on $\mathbf{t}$.) Then we can write $\mathbf{x} = \mu_0(\mathbf{t}) + R^T(\mathbf{t})\mathbf{y}$ and, as a consequence, $s_0(\mathbf{w}) = \|\mathbf{y}\|^2$. Note that $\mathbf{y}$ can be factorized as $\mathbf{y} = \|\mathbf{y}\|\,\mathbf{U}$, where $\|\mathbf{y}\|$ and $\mathbf{U}$ are independent and $\mathbf{U}$ has the uniform distribution on the unit sphere in $\mathbb{R}^m$, implying that $\mathrm{E}(\mathbf{U}) = \mathbf{0}$ and $\mathrm{E}(\mathbf{U}\mathbf{U}^T) = I_m/m$.

Since

$$\begin{aligned} \boldsymbol{\psi}_1(\mathbf{w}, \boldsymbol{\xi}^*) &= \left(\frac{\nu + m}{\nu\beta_0 + \|\mathbf{y}\|^2}\right) B^T(\mathbf{t})\Sigma_0^{-1}(\mathbf{t})R^T(\mathbf{t})\mathbf{y} \\ &= \left\{\frac{(\nu + m)\|\mathbf{y}\|}{\nu\beta_0 + \|\mathbf{y}\|^2}\right\} B^T(\mathbf{t})R^{-1}(\mathbf{t})\mathbf{U}, \end{aligned}$$

it is clear that $\mathrm{E}_0\{\boldsymbol{\psi}_1(\mathbf{w}, \boldsymbol{\xi}^*)\} = \mathbf{0}$ regardless of the value of β_0. Note that

$$\begin{aligned} \mathrm{E}\left\{\frac{(\nu + m)\|\mathbf{y}\|}{\nu\beta_0 + \|\mathbf{y}\|^2}\right\} &\leq \mathrm{E}^{\frac{1}{2}}\left(\frac{\nu + m}{\nu\beta_0 + \|\mathbf{y}\|^2}\right)\mathrm{E}^{\frac{1}{2}}\left\{\frac{(\nu + m)\|\mathbf{y}\|^2}{\nu\beta_0 + \|\mathbf{y}\|^2}\right\} \\ &\leq \left(\frac{\nu + m}{\nu\beta_0}\right)^{\frac{1}{2}} (\nu + m)^{\frac{1}{2}}, \end{aligned}$$

which is finite even if $\|\mathbf{y}\|$ does not have finite moments of any order.

Similarly,

$$\begin{aligned} S_0(\mathbf{w}) &= -B^T(\mathbf{t})\Sigma_0^{-1}(\mathbf{t})B(\mathbf{t}) \\ &\quad + \left\{ \frac{(\nu+m)\|\mathbf{y}\|^2}{\beta_0\nu + \|\mathbf{y}\|^2} \right\} B^T(\mathbf{t})R^{-1}(\mathbf{t})\mathbf{U}\mathbf{U}^T\{R^{-1}(\mathbf{t})\}^T B(\mathbf{t}) \end{aligned}$$

and then

$$\mathrm{E}_0\{S_0(\mathbf{w})\} = \left[-1 + \mathrm{E}\left\{ \frac{(\nu+m)\|\mathbf{y}\|^2}{\beta_0\nu + \|\mathbf{y}\|^2} \right\} \frac{1}{m} \right] \mathrm{E}\{B^T(\mathbf{t})\Sigma_0^{-1}(\mathbf{t})B(\mathbf{t})\},$$

which is going to be zero as long as β_0 satisfies

$$\mathrm{E}\left\{ \frac{(\nu+m)\|\mathbf{y}\|^2}{\beta_0\nu + \|\mathbf{y}\|^2} \right\} = m. \tag{4}$$

The fact that $\mathrm{E}_0\{S_0(\mathbf{w})\} = 0$ clearly implies $\mathrm{E}_0\{\boldsymbol{\psi}_2(\mathbf{w},\boldsymbol{\xi}^*)\} = \mathbf{0}$ and $\mathrm{E}_0\{\boldsymbol{\psi}_3(\mathbf{w},\boldsymbol{\xi}^*)\} = \mathbf{0}$. Note that (4) always has a positive solution β_0, since

$$\lim_{\beta \searrow 0} \mathrm{E}\left\{ \frac{(\nu+m)\|\mathbf{y}\|^2}{\beta\nu + \|\mathbf{y}\|^2} \right\} = \nu + m$$

and

$$\lim_{\beta \nearrow \infty} \mathrm{E}\left\{ \frac{(\nu+m)\|\mathbf{y}\|^2}{\beta\nu + \|\mathbf{y}\|^2} \right\} = 0.$$

Again, note that

$$\mathrm{E}\left\{ \frac{(\nu+m)\|\mathbf{y}\|^2}{\beta_0\nu + \|\mathbf{y}\|^2} \right\} \leq \nu + m,$$

so the expectation is finite even if $\|\mathbf{y}\|$ does note have finite moments of any order.

Finally,

$$\begin{aligned}
\psi_4(\mathbf{w}, \boldsymbol{\xi}^*) &= -\frac{1}{2}\frac{1}{\beta_0}\mathrm{tr}\{\Sigma_0^{-1}(\mathbf{t})\} \\
&\quad +\frac{1}{2}\left(\frac{\nu+m}{\nu+\|\mathbf{y}\|^2/\beta_0}\right)\frac{1}{\beta_0^2}\mathbf{y}^T R(\mathbf{t})\Sigma_0^{-2}(\mathbf{t})R^T(\mathbf{t})\mathbf{y} \\
&= -\frac{1}{2}\frac{1}{\beta_0}\mathrm{tr}\{\Sigma_0^{-1}(\mathbf{t})\} \\
&\quad +\frac{1}{2}\left\{\frac{(\nu+m)\|\mathbf{y}\|^2}{\beta_0\nu+\|\mathbf{y}\|^2}\right\}\frac{1}{\beta_0}\mathbf{U}^T\{R^T(\mathbf{t})\}^{-1}R^{-1}(\mathbf{t})\mathbf{U}.
\end{aligned}$$

Since

$$\mathbf{U}^T\{R^T(\mathbf{t})\}^{-1}R^{-1}(\mathbf{t})\mathbf{U} = \mathrm{tr}[R^{-1}(\mathbf{t})\mathbf{U}\mathbf{U}^T\{R^T(\mathbf{t})\}^{-1}],$$

we have

$$\begin{aligned}
\mathrm{E}_0[\mathbf{U}^T\{R^T(\mathbf{t})\}^{-1}R^{-1}(\mathbf{t})\mathbf{U}] &= \frac{1}{m}\mathrm{E}_0(\mathrm{tr}[R^{-1}(\mathbf{t})\{R^T(\mathbf{t})\}^{-1}]) \\
&= \frac{1}{m}\mathrm{E}_0[\mathrm{tr}\{\Sigma_0^{-1}(\mathbf{t})\}]
\end{aligned}$$

and then

$$\mathrm{E}_0\{\psi_4(\mathbf{w}, \boldsymbol{\xi}^*)\} = \frac{1}{2\beta_0}\mathrm{E}_0[\mathrm{tr}\{\Sigma_0^{-1}(\mathbf{t})\}]\left[-1+\mathrm{E}\left\{\frac{(\nu+m)\|\mathbf{y}\|^2}{\beta_0\nu+\|\mathbf{y}\|^2}\right\}\frac{1}{m}\right]$$

which again is equal to zero by (4).

5.2. Proof of Theorem 2

Since $\boldsymbol{\psi}(\mathbf{w}, \boldsymbol{\xi})$ is (up to constants) the gradient of $\log f(\mathbf{x})$, $\partial\boldsymbol{\psi}/\partial\boldsymbol{\xi}^T$ is (up to constants) the Hessian of $\log f(\mathbf{x})$ and then M is symmetric. Concretely,

$$\frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{\xi}^T} = \begin{bmatrix}
\frac{\partial\boldsymbol{\psi}_1}{\partial\boldsymbol{\theta}^T} & * & * & * \\
\frac{\partial\boldsymbol{\psi}_2}{\partial\boldsymbol{\theta}^T} & \frac{\partial\boldsymbol{\psi}_2}{\partial\mathrm{vec}(H)^T} & * & * \\
\frac{\partial\boldsymbol{\psi}_3}{\partial\boldsymbol{\theta}^T} & \frac{\partial\boldsymbol{\psi}_3}{\partial\mathrm{vec}(H)^T} & \frac{\partial\boldsymbol{\psi}_3}{\partial\boldsymbol{\lambda}^T} & * \\
\frac{\partial\psi_4}{\partial\boldsymbol{\theta}^T} & \frac{\partial\psi_4}{\partial\mathrm{vec}(H)^T} & \frac{\partial\psi_4}{\partial\boldsymbol{\lambda}^T} & \frac{\partial\psi_4}{\partial\sigma^2}
\end{bmatrix} \tag{5}$$

and

$$M_{11} = \mathrm{E}_0 \left\{ \frac{\partial \boldsymbol{\psi}_1(\mathbf{w}, \boldsymbol{\xi}^*)}{\partial \boldsymbol{\theta}^T} \right\}.$$

To prove Theorem 2 we need to find M_{11} explicitly and show that the expectations of the other three blocks in the first column of (5) are zero. We split the proof into different subsections to make it easier to follow.

5.2.1. Block (1,1): M_{11}

Differentiation of $\boldsymbol{\psi}_1$ with respect to $\boldsymbol{\theta}$ gives

$$\begin{aligned} \mathrm{d}\boldsymbol{\psi}_1 &= -\frac{(\nu+m)}{\{\nu+s(\mathbf{w},\boldsymbol{\xi})\}^2}\{\mathrm{d}s(\mathbf{w},\boldsymbol{\xi})\}B^T(\mathbf{t})\Sigma^{-1}(\mathbf{t},\boldsymbol{\xi})\{\mathbf{x}-B(\mathbf{t})\boldsymbol{\theta}\} \\ &\quad -\left\{\frac{\nu+m}{\nu+s(\mathbf{w},\boldsymbol{\xi})}\right\}B^T(\mathbf{t})\Sigma^{-1}(\mathbf{t},\boldsymbol{\xi})B(\mathbf{t})\mathrm{d}\boldsymbol{\theta}, \end{aligned}$$

and

$$\mathrm{d}s(\mathbf{w},\boldsymbol{\xi}) = -2\{\mathbf{x}-B(\mathbf{t})\boldsymbol{\theta}\}^T\Sigma^{-1}(\mathbf{t},\boldsymbol{\xi})B(\mathbf{t})\mathrm{d}\boldsymbol{\theta}.$$

Therefore

$$\begin{aligned} \frac{\partial \boldsymbol{\psi}_1(\mathbf{w},\boldsymbol{\xi})}{\partial \boldsymbol{\theta}^T} &= \frac{2(\nu+m)}{\{\nu+s(\mathbf{w},\boldsymbol{\xi})\}^2}B^T(\mathbf{t})\Sigma^{-1}(\mathbf{t},\boldsymbol{\xi})\{\mathbf{x}-B(\mathbf{t})\boldsymbol{\theta}\}\{\mathbf{x}-B(\mathbf{t})\boldsymbol{\theta}\}^T\Sigma^{-1}(\mathbf{t},\boldsymbol{\xi})B(\mathbf{t}) \\ &\quad -\left\{\frac{\nu+m}{\nu+s(\mathbf{w},\boldsymbol{\xi})}\right\}B^T(\mathbf{t})\Sigma^{-1}(\mathbf{t},\boldsymbol{\xi})B(\mathbf{t}). \end{aligned}$$

Evaluating at $\boldsymbol{\xi} = \boldsymbol{\xi}^*$ and using the notation of the proof of Theorem 1, under F_0 we have

$$\begin{aligned} \frac{\partial \boldsymbol{\psi}_1(\mathbf{w},\boldsymbol{\xi}^*)}{\partial \boldsymbol{\theta}^T} &= \frac{2(\nu+m)}{(\nu+\|\mathbf{y}\|^2/\beta_0)^2}B^T(\mathbf{t})\frac{1}{\beta_0}\Sigma_0^{-1}(\mathbf{t})R^T(\mathbf{t})\mathbf{y}\mathbf{y}R(\mathbf{t})\frac{1}{\beta_0}\Sigma_0^{-1}(\mathbf{t})B(\mathbf{t}) \\ &\quad -\left(\frac{\nu+m}{\nu+\|\mathbf{y}\|^2/\beta_0}\right)B^T(\mathbf{t})\frac{1}{\beta_0}\Sigma_0^{-1}(\mathbf{t})B(\mathbf{t}) \\ &= \left\{\frac{2(\nu+m)\|\mathbf{y}\|^2/\beta_0}{(\nu+\|\mathbf{y}\|^2/\beta_0)^2}\right\}B^T(\mathbf{t})\frac{1}{\beta_0}\Sigma_0^{-1}(\mathbf{t})R^T(\mathbf{t})\mathbf{U}\mathbf{U}^TR(\mathbf{t})\Sigma_0^{-1}(\mathbf{t})B(\mathbf{t}) \\ &\quad -\left(\frac{\nu+m}{\nu+\|\mathbf{y}\|^2/\beta_0}\right)B^T(\mathbf{t})\frac{1}{\beta_0}\Sigma_0^{-1}(\mathbf{t})B(\mathbf{t}). \end{aligned}$$

Since $\mathrm{E}(\mathbf{U}\mathbf{U}^T) = m^{-1}I_m$ and $R^T(\mathbf{t})R(\mathbf{t}) = \Sigma_0(\mathbf{t})$, the expression given for M_{11} in the statement of Theorem 2 follows (noting that $s_0(\mathbf{w}) = \|\mathbf{y}\|^2$ under F_0.) It is easy to see that M_{11} is invertible (in fact, negative definite) because

$$\begin{aligned} & \frac{2}{m}\left\{\frac{(\nu+m)\|\mathbf{y}\|^2/\beta_0}{(\nu+\|\mathbf{y}\|^2/\beta_0)^2}\right\} - \left(\frac{\nu+m}{\nu+\|\mathbf{y}\|^2/\beta_0^2}\right) \\ = & \frac{(\nu+m)\{(2/m-1)\|\mathbf{y}\|^2/\beta_0-\nu\}}{(\nu+\|\mathbf{y}\|^2/\beta_0)^2} < 0. \end{aligned}$$

Block (2,1)

Since

$$\begin{aligned} \boldsymbol{\psi}_2(\mathbf{w},\boldsymbol{\xi}) &= \mathrm{vec}\{(I_d - JHH^T)S(\mathbf{w},\boldsymbol{\xi})H\} \\ &= \{H^T \otimes (I_d - JHH^T)\}\mathrm{vec}\{S(\mathbf{w},\boldsymbol{\xi})\}, \end{aligned}$$

[using the property $\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B)$,] it is clear that

$$\mathrm{E}_0\left\{\frac{\partial\boldsymbol{\psi}_2(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\boldsymbol{\theta}^T}\right\} = 0 \Leftrightarrow \mathrm{E}_0\left[\frac{\partial\mathrm{vec}\{S(\mathbf{w},\boldsymbol{\xi}^*)\}}{\partial\boldsymbol{\theta}^T}\right] = 0,$$

so we will prove the latter. Differentiating with respect to $\boldsymbol{\theta}$ we get

$$\begin{aligned} \mathrm{d}S &= -\frac{\nu+m}{(\nu+s)^2}\,\mathrm{d}s\; B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B \\ &\quad -\left(\frac{\nu+m}{\nu+s}\right) B^T\Sigma^{-1}\,\mathrm{d}\boldsymbol{\mu}\,(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B \\ &\quad -\left(\frac{\nu+m}{\nu+s}\right) B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\,(\mathrm{d}\boldsymbol{\mu})^T\Sigma^{-1}B, \end{aligned}$$

with $\mathrm{d}s = -2(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}\mathrm{d}\boldsymbol{\mu}$ and $\mathrm{d}\boldsymbol{\mu} = B\mathrm{d}\boldsymbol{\theta}$. Then we can write

$$\begin{aligned} \mathrm{d}S &= \frac{2(\nu+m)}{(\nu+s)^2}\,B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B\,\mathrm{d}\boldsymbol{\theta}\,(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B \qquad (6) \\ &\quad -\left(\frac{\nu+m}{\nu+s}\right) B^T\Sigma^{-1}B\,\mathrm{d}\boldsymbol{\theta}\,(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B \\ &\quad -\left(\frac{\nu+m}{\nu+s}\right) B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathrm{d}\boldsymbol{\theta})^TB^T\Sigma^{-1}B, \end{aligned}$$

so

$$
\begin{aligned}
\operatorname{vec}(\mathrm{d}S) \;=\; & \left\{\frac{2(\nu+m)}{(\nu+s)^2}\right\}\{B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\otimes B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B\}\ \mathrm{d}\boldsymbol{\theta} \\
& -\left(\frac{\nu+m}{\nu+s}\right)\{B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\otimes B^T\Sigma^{-1}B\}\ \mathrm{d}\boldsymbol{\theta} \\
& -\left(\frac{\nu+m}{\nu+s}\right)\{B^T\Sigma^{-1}B\otimes B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}\ \mathrm{d}\boldsymbol{\theta}
\end{aligned}
$$

(in the last line we used that $\operatorname{vec}\{(\mathrm{d}\boldsymbol{\theta})^T\} = \mathrm{d}\boldsymbol{\theta}$, since $\mathrm{d}\boldsymbol{\theta}$ is a vector.)

Then, using the notation of the proof of Theorem 1, under F_0 we have

$$
\frac{\partial \operatorname{vec}\{S(\mathbf{w},\boldsymbol{\xi}^*)\}}{\partial\boldsymbol{\theta}^T} =
$$

$$
\begin{aligned}
& \frac{1}{\beta_0^3}\left\{\frac{2(\nu+m)\left\|\mathbf{y}\right\|^3}{(\nu+\left\|\mathbf{y}\right\|^2/\beta_0)^2}\right\}(B^T\Sigma_0^{-1}\mathbf{U}\otimes B^T\Sigma_0^{-1}\mathbf{U}\mathbf{U}^T\Sigma_0^{-1}B) \\
& \quad -\frac{1}{\beta_0^2}\left\{\frac{(\nu+m)\left\|\mathbf{y}\right\|}{\nu+\left\|\mathbf{y}\right\|^2/\beta_0}\right\}(B^T\Sigma_0^{-1}\mathbf{U}\otimes B^T\Sigma_0^{-1}B) \\
& \quad -\frac{1}{\beta_0^2}\left\{\frac{(\nu+m)\left\|\mathbf{y}\right\|}{\nu+\left\|\mathbf{y}\right\|^2/\beta_0}\right\}(B^T\Sigma_0^{-1}B\otimes B^T\Sigma_0^{-1}\mathbf{U}),
\end{aligned}
$$

where the dependence on $\mathbf{t}$ has been omitted for ease of notation. It is clear that the last two terms have zero expectation because $\mathrm{E}(\mathbf{U}) = \mathbf{0}$. As for the first term, note that

$$
\mathrm{E}(B^T\Sigma_0^{-1}\mathbf{U}\otimes B^T\Sigma_0^{-1}\mathbf{U}\mathbf{U}^T\Sigma_0^{-1}B\ |\mathbf{t}) = B^T\Sigma_0^{-1}\mathrm{E}(\mathbf{U}\otimes\mathbf{U}\mathbf{U}^T\Sigma_0^{-1}B\ |\mathbf{t})
$$

which is zero because $\mathrm{E}(U_k\mathbf{U}\mathbf{U}^T) = 0$ for all $k = 1,\ldots,m$.

5.2.2. Block (3,1)

From the preceding discussion it is immediate that this block is zero, since

$$
\psi_{3k}(\mathbf{w},\boldsymbol{\xi}) = \boldsymbol{\eta}_k^T S(\mathbf{w},\boldsymbol{\xi})\boldsymbol{\eta}_k = (\boldsymbol{\eta}_k^T\otimes\boldsymbol{\eta}_k^T)\operatorname{vec}\{S(\mathbf{w},\boldsymbol{\xi})\}
$$

and then

$$\mathrm{E}_0\left\{\frac{\partial\psi_{3k}(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\boldsymbol{\theta}^T}\right\} = (\boldsymbol{\eta}_k^T\otimes\boldsymbol{\eta}_k^T)\mathrm{E}_0\left[\frac{\partial\mathrm{vec}\{S(\mathbf{w},\boldsymbol{\xi}^*)\}}{\partial\boldsymbol{\theta}^T}\right] = 0.$$

Block (4,1)

Differentiating with respect to $\boldsymbol{\theta}$ we see that

$$\begin{aligned}\mathrm{d}\psi_4 &= -\frac{1}{2}\left\{\frac{\nu+m}{(\nu+s)^2}\right\}\mathrm{d}s\;(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-2}(\mathbf{x}-\boldsymbol{\mu}) \\ &\quad -\frac{1}{2}\left(\frac{\nu+m}{\nu+s}\right)(\mathrm{d}\boldsymbol{\mu})^T\Sigma^{-2}(\mathbf{x}-\boldsymbol{\mu}) \\ &\quad -\frac{1}{2}\left(\frac{\nu+m}{\nu+s}\right)(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-2}\,\mathrm{d}\boldsymbol{\mu},\end{aligned}$$

and since $\mathrm{d}s$ involves a factor $\mathbf{x}-\boldsymbol{\mu}$, it is clear that $\mathrm{d}\psi_4$ follows the same pattern as $\mathrm{d}S$ in (6): the first term involves the factor $\mathbf{x}-\boldsymbol{\mu}$ three times and the last two terms involve the factor $\mathbf{x}-\boldsymbol{\mu}$ once each, so the three terms will be zero under F_0 when $\boldsymbol{\xi}=\boldsymbol{\xi}^*$.

5.2.3. Block (2,2)

Differentiating with respect to H we have

$$\begin{aligned}\mathrm{d}\boldsymbol{\psi}_2(\mathbf{w},\boldsymbol{\xi}) &= -\mathrm{vec}\{J(\mathrm{d}H)H^TS(\mathbf{w},\boldsymbol{\xi})H\} \\ &\quad -\mathrm{vec}\{JH(\mathrm{d}H^T)S(\mathbf{w},\boldsymbol{\xi})H\} \\ &\quad +\mathrm{vec}\{(I_p-JHH^T)\mathrm{d}S(\mathbf{w},\boldsymbol{\xi})H\} \\ &\quad +\mathrm{vec}\{(I_p-JHH^T)S(\mathbf{w},\boldsymbol{\xi})\mathrm{d}H\}.\end{aligned}$$

When proving Theorem 1 we saw that $\mathrm{E}_0\{S(\mathbf{w},\boldsymbol{\xi}^*)\}=0$, so only the third term of the preceding display will be non-zero for $\boldsymbol{\xi}=\boldsymbol{\xi}^*$ when expectations under F_0 are taken.

Differentiating S with respect to H we obtain

$$\begin{aligned} \mathrm{d}S &= B^T\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}B \\ &\quad - \left\{\frac{\nu+m}{(\nu+s)^2}\right\} \mathrm{d}s\; B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B \\ &\quad - \left(\frac{\nu+m}{\nu+s}\right) B^T\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}B \\ &\quad - \left(\frac{\nu+m}{\nu+s}\right) B^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}B, \end{aligned}$$

where

$$\mathrm{d}s = -(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}),$$

and

$$\mathrm{d}\Sigma = B(\mathrm{d}H)\Lambda H^T B^T + BH\Lambda(\mathrm{d}H)^T B^T.$$

To simplify notation let us call $\Omega = \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}$. Then

$$\begin{aligned} \mathrm{d}S &= B^T\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}B \\ &\quad + \left\{\frac{\nu+m}{(\nu+s)^2}\right\} B^T\Omega(\mathrm{d}\Sigma)\Omega B \\ &\quad - \left(\frac{\nu+m}{\nu+s}\right) B^T\Sigma^{-1}(\mathrm{d}\Sigma)\Omega B \\ &\quad - \left(\frac{\nu+m}{\nu+s}\right) B^T\Omega(\mathrm{d}\Sigma)\Sigma^{-1}B. \end{aligned}$$

At $\boldsymbol{\xi} = \boldsymbol{\xi}^*$ and under model F_0 the matrix Ω comes down to

$$\Omega_0 = \frac{1}{\beta_0^2}R^{-1}\mathbf{y}\mathbf{y}^T(R^{-1})^T$$

and then

$$\mathrm{E}_0(\Omega_0|\,\|\mathbf{y}\|\,,\mathbf{t}) = \|\mathbf{y}\|^2\,\frac{1}{\beta_0^2 m}\Sigma_0^{-1}. \tag{7}$$

Also, using that

$$\mathrm{E}(\mathbf{U}\mathbf{U}^T \otimes \mathbf{U}\mathbf{U}^T) = \frac{1}{m(m+2)}\{I_{m^2} + K_{mm} + \mathrm{vec}(I_m)\mathrm{vec}(I_m)^T\},$$

where K_{mm} is the commutation matrix (see e.g. proposition 13.2 of Bilodeau and

Brenner 1999,) we see that

$$\mathrm{E}_0\{\mathrm{vec}(\Omega_0(\mathrm{d}\Sigma)\Omega_0)|\,\|\mathbf{y}\|\,,\mathbf{t}\} = \mathrm{E}_0(\Omega_0 \otimes \Omega_0|\,\|\mathbf{y}\|\,,\mathbf{t})\mathrm{vec}(\mathrm{d}\Sigma)$$

$$\begin{aligned}
&= \frac{\|\mathbf{y}\|^4}{\beta_0^4}(R^{-1}\otimes R^{-1})\mathrm{E}(\mathbf{U}\mathbf{U}^T\otimes \mathbf{U}\mathbf{U}^T)\{(R^{-1})^T\otimes (R^{-1})^T\}\mathrm{vec}(\mathrm{d}\Sigma)\\
&= \frac{\|\mathbf{y}\|^4}{\beta_0^4 m(m+2)}(R^{-1}\otimes R^{-1})\{I_{m^2}+K_{mm}+\mathrm{vec}(I_m)\mathrm{vec}(I_m)^T\}\mathrm{vec}\{(R^{-1})^T\mathrm{d}\Sigma R^{-1}\}\\
&= \frac{\|\mathbf{y}\|^4}{\beta_0^4 m(m+2)}[\mathrm{vec}\{\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}\}+\mathrm{vec}\{\Sigma_0^{-1}(\mathrm{d}\Sigma)^T\Sigma_0^{-1}\}+\mathrm{vec}(\Sigma_0^{-1})\mathrm{tr}\{(\mathrm{d}\Sigma)\Sigma_0^{-1}\}].
\end{aligned}$$

Note that $\mathrm{d}\Sigma$ is symmetric in this case. Therefore, "unstacking" the preceding expression we obtain

$$\mathrm{E}_0\{\Omega_0(\mathrm{d}\Sigma)\Omega_0|\,\|\mathbf{y}\|\} = \frac{\|\mathbf{y}\|^4}{\beta_0^4 m(m+2)}[2\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}+\mathrm{tr}\{(\mathrm{d}\Sigma)\Sigma_0^{-1}\}\Sigma_0^{-1}]. \quad (8)$$

Putting everything together,

$$\mathrm{E}_0\{\mathrm{d}S(\mathbf{w},\boldsymbol{\xi}^*)\} =$$

$$\begin{aligned}
&\frac{1}{\beta_0^2}\mathrm{E}\{B^T\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}B\}\\
&+\frac{1}{\beta_0^4 m(m+2)}\mathrm{E}\left\{\frac{(\nu+m)\|\mathbf{y}\|^4}{(\nu+\|\mathbf{y}\|^2/\beta_0)^2}\right\}\mathrm{E}[2B^T\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}B+\mathrm{tr}\{(\mathrm{d}\Sigma)\Sigma_0^{-1}\}B^T\Sigma_0^{-1}B]\\
&-\frac{2}{\beta_0^3 m}\left\{\frac{(\nu+m)\|\mathbf{y}\|^2}{\nu+\|\mathbf{y}\|^2/\beta_0}\right\}\mathrm{E}\{B^T\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}B\}.
\end{aligned}$$

Using (4) we can simplify this further:

$$\begin{aligned}
\mathrm{E}_0\{\mathrm{d}S(\mathbf{w},\boldsymbol{\xi}^*)\} &= \left\{-\frac{1}{\beta_0^2}+\frac{2\zeta_0}{\beta_0^2 m(m+2)}\right\}\mathrm{E}\{B^T\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}B\} \quad (9)\\
&\quad+\frac{\zeta_0}{\beta_0^2 m(m+2)}\mathrm{E}[\mathrm{tr}\{(\mathrm{d}\Sigma)\Sigma_0^{-1}\}B^T\Sigma_0^{-1}B],
\end{aligned}$$

where

$$\zeta_0 = \mathrm{E}\left\{ \frac{(\nu + m)\left\|\mathbf{y}\right\|^4}{(\beta_0\nu + \left\|\mathbf{y}\right\|^2)^2} \right\}.$$

Then

$$\begin{aligned}
\mathrm{E}_0[\mathrm{vec}\{\mathrm{d}S(\mathbf{w}, \boldsymbol{\xi}^*)\}] &= \left\{-\frac{1}{\beta_0^2} + \frac{2\zeta_0}{\beta_0^2 m(m+2)}\right\} \mathrm{E}\{(B^T\Sigma_0^{-1} \otimes B^T\Sigma_0^{-1})\mathrm{vec}(\mathrm{d}\Sigma)\} \\
&\quad + \frac{\zeta_0}{\beta_0^2 m(m+2)} \mathrm{E}\{\mathrm{vec}(B^T\Sigma_0^{-1}B)\mathrm{vec}(\Sigma_0^{-1})^T\mathrm{vec}(\mathrm{d}\Sigma)\}.
\end{aligned}$$

(This expression will be used again many times, since only $\mathrm{d}\Sigma$ changes when differentiating with respect to $\boldsymbol{\lambda}$ or σ^2.) For the particular case we are dealing with, $\mathrm{d}\Sigma$ at $\boldsymbol{\xi} = \boldsymbol{\xi}^*$ it comes down to

$$\mathrm{d}\Sigma = \beta_0\{B(\mathrm{d}H)\Lambda_0 H_0^T B^T + BH_0\Lambda_0(\mathrm{d}H)^T B^T\}$$

and then

$$(\mathrm{d}\Sigma) = (\beta_0 BH_0\Lambda_0 \otimes B)\mathrm{vec}(\mathrm{d}H) + (\beta_0 B \otimes BH_0\Lambda_0)K_{pd}\mathrm{vec}(\mathrm{d}H).$$

Using the properties *(i)* $K_{tr}(A \otimes B)K_{su} = B \otimes A$ for any $A \in \mathbb{R}^{r\times s}$ and $B \in \mathbb{R}^{t\times u}$, and *(ii)* $K_{su}^T = K_{su}^{-1} = K_{us}$, we can write

$$\begin{aligned}
(\mathrm{d}\Sigma) &= \{(\beta_0 BH_0\Lambda_0 \otimes B) + K_{mm}(\beta_0 BH_0\Lambda_0 \otimes B)\}\mathrm{vec}(\mathrm{d}H) \\
&= \beta_0(I_{mm} + K_{mm})(BH_0\Lambda_0 \otimes B)\mathrm{vec}(\mathrm{d}H).
\end{aligned}$$

Since

$$\mathrm{E}_0\left\{\mathrm{d}\boldsymbol{\psi}_2(\mathbf{w}, \boldsymbol{\xi}^*)\right\} = \{H_0^T \otimes (I_p - JH_0H_0^T)\}\mathrm{vec}[\mathrm{E}_0\{\mathrm{d}S(\mathbf{w}, \boldsymbol{\xi}^*)\}]$$

we finally obtain

$$\begin{aligned}
\mathrm{E}_0\left\{\frac{\partial \boldsymbol{\psi}_2(\mathbf{w},\boldsymbol{\xi}^*)}{\partial \mathrm{vec}(H)^T}\right\} &= \frac{1}{\beta_0}\{H_0^T \otimes (I_p - JH_0H_0^T)\} \times \\
&\mathrm{E}[\{(-1+\frac{2\zeta_0}{m(m+2)})(B^T\Sigma_0^{-1} \otimes B^T\Sigma_0^{-1}) \\
&+\frac{\zeta_0}{m(m+2)}\mathrm{vec}(B^T\Sigma_0^{-1}B)\mathrm{vec}(\Sigma_0^{-1})^T\} \\
&\times(I_{mm}+K_{mm})(BH_0\Lambda_0 \otimes B)],
\end{aligned}$$

or equivalently,

$$\begin{aligned}
\mathrm{E}_0\left\{\frac{\partial \boldsymbol{\psi}_2(\mathbf{w},\boldsymbol{\xi}^*)}{\partial \mathrm{vec}(H)^T}\right\} &= \frac{1}{\beta_0}\mathrm{E}[\{(-1+\frac{2\zeta_0}{m(m+2)})(H_0^TB^T\Sigma_0^{-1} \otimes (I_p - JH_0H_0^T)B^T\Sigma_0^{-1}) \\
&+\frac{\zeta_0}{m(m+2)}\mathrm{vec}((I_p - JH_0H_0^T)B^T\Sigma_0^{-1}BH_0)\mathrm{vec}(\Sigma_0^{-1})^T\} \\
&\times(I_{mm}+K_{mm})(BH_0\Lambda_0 \otimes B)]
\end{aligned}$$

and also

$$\mathrm{E}_0\left\{\frac{\partial \boldsymbol{\psi}_2(\mathbf{w},\boldsymbol{\xi}^*)}{\partial \mathrm{vec}(H)^T}\right\} =$$

$$\begin{aligned}
&\frac{1}{\beta_0}\mathrm{E}\{(-1+\frac{2\zeta_0}{m(m+2)})(H_0^TB^T\Sigma_0^{-1}BH_0\Lambda_0 \otimes (I_p - JH_0H_0^T)B^T\Sigma_0^{-1}B) \\
&+(-1+\frac{2\zeta_0}{m(m+2)})(H_0^TB^T\Sigma_0^{-1}B \otimes (I_p - JH_0H_0^T)B^T\Sigma_0^{-1}BH_0\Lambda_0)K_{pd} \\
&+\frac{\zeta_0}{m(m+2)}\mathrm{vec}((I_p - JH_0H_0^T)B^T\Sigma_0^{-1}BH_0)\mathrm{vec}(B^T\Sigma_0^{-1}BH_0\Lambda_0)^T \\
&+\frac{\zeta_0}{m(m+2)}\mathrm{vec}((I_p - JH_0H_0^T)B^T\Sigma_0^{-1}BH_0)\mathrm{vec}(\Lambda_0H_0^TB^T\Sigma_0^{-1}B)^TK_{pd}\}.
\end{aligned}$$

5.2.4. Block (3,2)

To compute $\mathrm{d}\boldsymbol{\psi}_3$, note that we can write

$$\begin{aligned}\boldsymbol{\psi}_3(\mathbf{w},\boldsymbol{\xi}) &= \sum_{k=1}^{d}\mathbf{e}_k\boldsymbol{\eta}_k^T S(\mathbf{w},\boldsymbol{\xi})\boldsymbol{\eta}_k \\ &= \sum_{k=1}^{d}\mathbf{e}_k\mathbf{e}_k^T H^T S(\mathbf{w},\boldsymbol{\xi})H\mathbf{e}_k \\ &= (\sum_{k=1}^{d}\mathbf{e}_k^T\otimes\mathbf{e}_k\mathbf{e}_k^T)(H^T\otimes H^T)\mathrm{vec}\{S(\mathbf{w},\boldsymbol{\xi})\},\end{aligned}$$

with $\mathbf{e}_k\in\mathbb{R}^d$. Differentiating with respect to H,

$$\begin{aligned}\mathrm{d}\boldsymbol{\psi}_3(\mathbf{w},\boldsymbol{\xi}) &= (\sum_{k=1}^{d}\mathbf{e}_k^T\otimes\mathbf{e}_k\mathbf{e}_k^T)\mathrm{d}(H^T\otimes H^T)\mathrm{vec}\{S(\mathbf{w},\boldsymbol{\xi})\} \\ &\quad +(\sum_{k=1}^{d}\mathbf{e}_k^T\otimes\mathbf{e}_k\mathbf{e}_k^T)(H^T\otimes H^T)\mathrm{vec}\{\mathrm{d}S(\mathbf{w},\boldsymbol{\xi})\}.\end{aligned}$$

The first term on the right hand side will vanish when expectations are taken under F_0 at $\boldsymbol{\xi}=\boldsymbol{\xi}^*$, because $\mathrm{E}_0\{S(\mathbf{w},\boldsymbol{\xi}^*)\}$. Then

$$\mathrm{E}_0\{\mathrm{d}\boldsymbol{\psi}_3(\mathbf{w},\boldsymbol{\xi})\} = (\sum_{k=1}^{d}\mathbf{e}_k^T\otimes\mathbf{e}_k\mathbf{e}_k^T)(H_0^T\otimes H_0^T)\mathrm{vec}[\mathrm{E}_0\{\mathrm{d}S(\mathbf{w},\boldsymbol{\xi}^*)\}],$$

where $\mathrm{E}_0\{\mathrm{d}S(\mathbf{w},\boldsymbol{\xi}^*)\}$ is exactly as in (9), including $\mathrm{d}\Sigma$, since we are still differentiating with respect to H. Then

$$\begin{aligned}\mathrm{E}_0\left\{\frac{\partial\boldsymbol{\psi}_3(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\mathrm{vec}(H)^T}\right\} &= \frac{1}{\beta_0}(\sum_{k=1}^{d}\mathbf{e}_k^T\otimes\mathbf{e}_k\mathbf{e}_k^T)(H_0^T\otimes H_0^T)\times \\ &\quad \mathrm{E}[\{(-1+\frac{2\zeta_0}{m(m+2)})(B^T\Sigma_0^{-1}\otimes B^T\Sigma_0^{-1}) \\ &\quad +\frac{\zeta_0}{m(m+2)}\mathrm{vec}(B^T\Sigma_0^{-1}B)\mathrm{vec}(\Sigma_0^{-1})^T\} \\ &\quad \times(I_{mm}+K_{mm})(BH_0\Lambda_0\otimes B)],\end{aligned}$$

or, more explicitly,

$$\mathrm{E}_0\left\{\frac{\partial \boldsymbol{\psi}_3(\mathbf{w},\boldsymbol{\xi}^*)}{\partial \mathrm{vec}(H)^T}\right\} = \frac{1}{\beta_0}(\sum_{k=1}^{d}\mathbf{e}_k^T\otimes\mathbf{e}_k\mathbf{e}_k^T)\times$$

$$\begin{aligned}
&\mathrm{E}\{(-1+\frac{2\zeta_0}{m(m+2)})(H_0^TB^T\Sigma_0^{-1}BH_0\Lambda_0\otimes H_0^TB^T\Sigma_0^{-1}B)\\
&+(-1+\frac{2\zeta_0}{m(m+2)})(H_0^TB^T\Sigma_0^{-1}B\otimes H_0^TB^T\Sigma_0^{-1}BH_0\Lambda_0)K_{pd}\\
&+\frac{\zeta_0}{m(m+2)}\mathrm{vec}(H_0^TB^T\Sigma_0^{-1}BH_0)\mathrm{vec}(B^T\Sigma_0^{-1}BH_0\Lambda_0)^T\\
&+\frac{\zeta_0}{m(m+2)}\mathrm{vec}(H_0^TB^T\Sigma_0^{-1}BH_0)\mathrm{vec}(\Lambda_0H_0^TB^T\Sigma_0^{-1}B)^TK_{pd}\}.
\end{aligned}$$

5.2.5. Block (4,2)

Differentiating ψ_4 with respect to H we get

$$\begin{aligned}
\mathrm{d}\psi_4 &= \frac{1}{2}\mathrm{tr}\{\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}\}\\
&\quad-\frac{1}{2}\left\{\frac{\nu+m}{(\nu+s)^2}\right\}(\mathrm{d}s)(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-2}(\mathbf{x}-\boldsymbol{\mu})\\
&\quad+\frac{1}{2}\left(\frac{\nu+m}{\nu+s}\right)(\mathbf{x}-\boldsymbol{\mu})^T(\mathrm{d}\Sigma^{-2})(\mathbf{x}-\boldsymbol{\mu}),
\end{aligned}$$

where

$$\mathrm{d}s = -(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

and

$$\begin{aligned}
\mathrm{d}\Sigma^{-2} &= (\mathrm{d}\Sigma^{-1})\Sigma^{-1}+\Sigma^{-1}(\mathrm{d}\Sigma^{-1})\\
&= -\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-2}+\Sigma^{-2}(\mathrm{d}\Sigma)\Sigma^{-1}.
\end{aligned}$$

Taking traces and rearranging factors, we obtain

$$\mathrm{d}\psi_4 = \frac{1}{2}\mathrm{tr}\{\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}\}+\frac{1}{2}\left\{\frac{\nu+m}{(\nu+s)^2}\right\}\mathrm{tr}\{\Omega(\mathrm{d}\Sigma)\Omega\}$$

$$-\frac{1}{2}\left(\frac{\nu+m}{\nu+s}\right)\mathrm{tr}\{(\mathrm{d}\Sigma)\Sigma^{-1}\Omega\}-\frac{1}{2}\left(\frac{\nu+m}{\nu+s}\right)\mathrm{tr}\{\Omega\Sigma^{-1}(\mathrm{d}\Sigma)\}.$$

Using (7) and (8) we get

$$\begin{aligned}
\mathrm{E}_0\{\mathrm{d}\psi_4(\mathbf{w},\boldsymbol{\xi}^*)\} &= \frac{1}{2\beta_0^2}\mathrm{E}[\mathrm{tr}\{\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}\}] \\
&+\frac{1}{2\beta_0^4 m(m+2)}\mathrm{E}\left\{\frac{(\nu+m)\left\|\mathbf{y}\right\|^4}{(\nu+\left\|\mathbf{y}\right\|^2/\beta_0)^2}\right\}\times \\
&\mathrm{E}[2\mathrm{tr}\{\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}\}+\mathrm{tr}\{(\mathrm{d}\Sigma)\Sigma_0^{-1}\}\mathrm{tr}(\Sigma_0^{-1})] \\
&-\frac{1}{\beta_0^3 m}\mathrm{E}\left\{\frac{(\nu+m)\left\|\mathbf{y}\right\|^2}{\nu+\left\|\mathbf{y}\right\|^2/\beta_0}\right\}\mathrm{E}[\mathrm{tr}\{\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}\}].
\end{aligned}$$

Using (4) we can simplify this expression to

$$\mathrm{E}_0\{\mathrm{d}\psi_4(\mathbf{w},\boldsymbol{\xi}^*)\} = -\frac{1}{2\beta_0^2}\mathrm{E}[\mathrm{tr}\{\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}\}] \tag{10}$$

$$+\frac{\zeta_0}{2\beta_0^2 m(m+2)}\mathrm{E}[2\mathrm{tr}\{\Sigma_0^{-1}(\mathrm{d}\Sigma)\Sigma_0^{-1}\}+\mathrm{tr}\{(\mathrm{d}\Sigma)\Sigma_0^{-1}\}\mathrm{tr}(\Sigma_0^{-1})]$$

$$\begin{aligned}
&= \frac{1}{2\beta_0^2}\mathrm{E}[\{-1+\frac{2\zeta_0}{m(m+2)}\}\mathrm{vec}(\Sigma_0^{-2})^T\mathrm{vec}(\mathrm{d}\Sigma) \\
&+\frac{\zeta_0}{m(m+2)}\mathrm{tr}(\Sigma_0^{-1})\mathrm{vec}(\Sigma_0^{-1})^T\mathrm{vec}(\mathrm{d}\Sigma)]
\end{aligned}$$

and then

$$\begin{aligned}
\mathrm{E}_0\left\{\frac{\partial\psi_4(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\mathrm{vec}(H)^T}\right\} &= \frac{1}{2\beta_0}\mathrm{E}[\{(-1+\frac{2\zeta_0}{m(m+2)})\mathrm{vec}(\Sigma_0^{-2})^T \\
&+\frac{\zeta_0}{m(m+2)}\mathrm{tr}(\Sigma_0^{-1})\mathrm{vec}(\Sigma_0^{-1})^T\}\times \\
&(I_{mm}+K_{mm})(BH_0\Lambda_0\otimes B)].
\end{aligned}$$

Now, since Σ_0 is symmetric, $\text{vec}(\Sigma_0^{-2})^T K_{mm} = \text{vec}(\Sigma_0^{-2})^T$ and similarly for Σ_0^{-1}. Then

$$\begin{aligned} \mathrm{E}_0\left\{\frac{\partial\psi_4(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\text{vec}(H)^T}\right\} &= \frac{1}{\beta_0}\mathrm{E}\{(-1+\frac{2\zeta_0}{m(m+2)})\text{vec}(B^T\Sigma_0^{-2}BH_0\Lambda_0)^T \\ &\quad +\frac{\zeta_0}{m(m+2)}\text{tr}(\Sigma_0^{-1})\text{vec}(B^T\Sigma_0^{-1}BH_0\Lambda_0)^T\}. \end{aligned}$$

5.2.6. Block (3,3)

Since $\psi_{3k}(\mathbf{w},\boldsymbol{\xi}) = \boldsymbol{\eta}_k^T S(\mathbf{w},\boldsymbol{\xi})\boldsymbol{\eta}_k$, differentiating with respect to $\boldsymbol{\lambda}$ we obtain

$$\mathrm{E}_0\{\mathrm{d}\boldsymbol{\psi}_3(\mathbf{w},\boldsymbol{\xi}^*)\} = \sum_{k=1}^{d}\mathbf{e}_k\boldsymbol{\eta}_{0k}^T\mathrm{E}_0\{\mathrm{d}S(\mathbf{w},\boldsymbol{\xi}^*)\}\boldsymbol{\eta}_{0k}$$

with $\mathrm{E}_0\{\mathrm{d}S(\mathbf{w},\boldsymbol{\xi}^*)\}$ as in (9), except that

$$\mathrm{d}\Sigma = \sum_{k=1}^{d}(\mathrm{d}\lambda_k)B\boldsymbol{\eta}_{0k}\boldsymbol{\eta}_{0k}^T B^T.$$

Then

$$\mathrm{E}_0\left\{\frac{\partial\psi_{3k}(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\lambda_j}\right\} =$$

$$\begin{aligned} &\left\{-\frac{1}{\beta_0^2}+\frac{2\zeta_0}{\beta_0^2 m(m+2)}\right\}\mathrm{E}\{\boldsymbol{\eta}_{0k}^T B^T\Sigma_0^{-1}B\boldsymbol{\eta}_{0j}\boldsymbol{\eta}_{0j}^T B^T\Sigma_0^{-1}B\boldsymbol{\eta}_{0k}\} \\ &+\frac{\zeta_0}{\beta_0^2 m(m+2)}\mathrm{E}\{\text{tr}(B\boldsymbol{\eta}_{0j}\boldsymbol{\eta}_{0j}^T B^T\Sigma_0^{-1})\boldsymbol{\eta}_{0k}^T B^T\Sigma_0^{-1}B\boldsymbol{\eta}_{0k}\}. \end{aligned}$$

Since

$$\begin{aligned} &\boldsymbol{\eta}_{0k}^T B^T\Sigma_0^{-1}B\boldsymbol{\eta}_{0j}\boldsymbol{\eta}_{0j}^T B^T\Sigma_0^{-1}B\boldsymbol{\eta}_{0k} \\ =\ &\mathbf{e}_k^T H_0^T B^T\Sigma_0^{-1}BH_0\mathbf{e}_j\mathbf{e}_j^T H_0^T B^T\Sigma_0^{-1}BH_0\mathbf{e}_k \\ =\ &\text{tr}(\mathbf{e}_k\mathbf{e}_k^T H_0^T B^T\Sigma_0^{-1}BH_0\mathbf{e}_j\mathbf{e}_j^T H_0^T B^T\Sigma_0^{-1}BH_0) \\ =\ &\text{vec}(\mathbf{e}_k\mathbf{e}_k^T)^T(H_0^T B^T\Sigma_0^{-1}BH_0\otimes H_0^T B^T\Sigma_0^{-1}BH_0)\text{vec}(\mathbf{e}_j\mathbf{e}_j^T), \end{aligned}$$

[where we used the property $\mathrm{tr}(ABCD) = \mathrm{vec}(A^T)^T(D^T \otimes B)\mathrm{vec}(C)$,]

$$
\begin{aligned}
\mathrm{tr}(B\boldsymbol{\eta}_{0j}\boldsymbol{\eta}_{0j}^T B^T \Sigma_0^{-1}) &= \mathrm{tr}(H_0^T B^T \Sigma_0^{-1} B H_0 \mathbf{e}_j \mathbf{e}_j^T) \\
&= \mathrm{vec}(H_0^T B^T \Sigma_0^{-1} B H_0)^T \mathrm{vec}(\mathbf{e}_j \mathbf{e}_j^T),
\end{aligned}
$$

and

$$
\begin{aligned}
\boldsymbol{\eta}_{0k}^T B^T \Sigma_0^{-1} B \boldsymbol{\eta}_{0k} &= \mathrm{tr}(\mathbf{e}_k \mathbf{e}_k^T H_0^T B^T \Sigma_0^{-1} B H_0) \\
&= \mathrm{vec}(\mathbf{e}_k \mathbf{e}_k^T)^T \mathrm{vec}(H_0^T B^T \Sigma_0^{-1} B H_0),
\end{aligned}
$$

we have

$$
\mathrm{E}_0\left\{\frac{\partial \psi_{3k}(\mathbf{w}, \boldsymbol{\xi}^*)}{\partial \lambda_j}\right\} = \mathrm{vec}(\mathbf{e}_k \mathbf{e}_k^T)^T \Upsilon \mathrm{vec}(\mathbf{e}_j \mathbf{e}_j^T),
$$

with

$$
\begin{aligned}
\Upsilon = {} & \left\{-\frac{1}{\beta_0^2} + \frac{2\zeta_0}{\beta_0^2 m(m+2)}\right\} \mathrm{E}(H_0^T B^T \Sigma_0^{-1} B H_0 \otimes H_0^T B^T \Sigma_0^{-1} B H_0) \\
& + \frac{\zeta_0}{\beta_0^2 m(m+2)} \mathrm{E}\{\mathrm{vec}(H_0^T B^T \Sigma_0^{-1} B H_0) \mathrm{vec}(H_0^T B^T \Sigma_0^{-1} B H_0)^T\}.
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathrm{E}_0\left\{\frac{\partial \boldsymbol{\psi}_3(\mathbf{w}, \boldsymbol{\xi}^*)}{\partial \boldsymbol{\lambda}^T}\right\} &= \sum_{k=1}^d \sum_{j=1}^d \mathrm{E}_0\left\{\frac{\partial \psi_{3k}(\mathbf{w}, \boldsymbol{\xi}^*)}{\partial \lambda_j}\right\} \mathbf{e}_k \mathbf{e}_j^T \\
&= \{\sum_{k=1}^d \mathbf{e}_k \mathrm{vec}(\mathbf{e}_k \mathbf{e}_k^T)^T\} \Upsilon \{\sum_{j=1}^d \mathrm{vec}(\mathbf{e}_j \mathbf{e}_j^T) \mathbf{e}_j^T\} \\
&= (\sum_{j=1}^d \mathbf{e}_j \mathbf{e}_j^T \otimes \mathbf{e}_j)^T \Upsilon (\sum_{j=1}^d \mathbf{e}_j \mathbf{e}_j^T \otimes \mathbf{e}_j).
\end{aligned}
$$

5.2.7. Block (4,3)

It is clear that we can use expression (10) for this block, too; the only difference is that $\mathrm{d}\Sigma$ is as in block (3,3). Then

$$\begin{aligned}\mathrm{E}_0\left\{\frac{\partial\psi_4(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\lambda_j}\right\} &= \left\{-\frac{1}{2\beta_0^2}+\frac{2\zeta_0}{2\beta_0^2 m(m+2)}\right\}\mathrm{E}(\boldsymbol{\eta}_{0j}^T B^T\Sigma_0^{-2}B\boldsymbol{\eta}_{0j})\\ &\quad+\frac{\zeta_0}{2\beta_0^2 m(m+2)}\mathrm{E}\{(\boldsymbol{\eta}_{0j}^T B^T\Sigma_0^{-1}B\boldsymbol{\eta}_{0j})\mathrm{tr}(\Sigma_0^{-1})\}.\end{aligned}$$

As above, we have

$$\boldsymbol{\eta}_{0j}^T B^T\Sigma_0^{-2}B\boldsymbol{\eta}_{0j}=\mathrm{vec}(H_0^T B^T\Sigma_0^{-2}BH_0)^T\mathrm{vec}(\mathbf{e}_j\mathbf{e}_j^T)$$

and similarly for the second term, so

$$\begin{aligned}\mathrm{E}_0\left\{\frac{\partial\psi_4(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\boldsymbol{\lambda}^T}\right\} &= \frac{1}{2\beta_0^2}\mathrm{E}\{(-1+\frac{2\zeta_0}{m(m+2)})\mathrm{vec}(H_0^T B^T\Sigma_0^{-2}BH_0)^T\\ &\quad+\frac{\zeta_0}{m(m+2)}\mathrm{tr}(\Sigma_0^{-1})\mathrm{vec}(H_0^T B^T\Sigma_0^{-1}BH_0)^T\}(\sum_{j=1}^{d}\mathbf{e}_j\mathbf{e}_j^T\otimes\mathbf{e}_j).\end{aligned}$$

5.2.8. Block (4,4)

Finally, using expression (10) once more, this time with $\mathrm{d}\Sigma=(\mathrm{d}\sigma^2)I_m$, we see that

$$\mathrm{E}_0\left\{\frac{\partial\psi_4(\mathbf{w},\boldsymbol{\xi}^*)}{\partial\sigma^2}\right\}=\frac{1}{2\beta_0^2}\mathrm{E}\{(-1+\frac{2\zeta_0}{m(m+2)})\mathrm{tr}(\Sigma_0^{-2})+\frac{\zeta_0}{m(m+2)}\mathrm{tr}(\Sigma_0^{-1})^2\}.$$

6. References

Bilodeau, M., and Brenner, D. (1999), *Theory of Multivariate Statistics*, New York: Springer.

Magnus, J. R., and Neudecker, H. (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics. Revised Edition*, New York: Wiley.