Supplement to "An Empirical Bayes Approach to Shrinkage Estimation on the Manifold of Symmetric Positive-Definite Matrices"

Chun-Hao Yang, Hani Doss and Baba C. Vemuri

Abstract

This document provides supporting material for the main paper, specifically the derivation of the risk function and the SURE, the proofs of Theorem 3 and 4, the implementation details for minimizing the SURE, and remarks regarding the resampling scheme in Section 4.2.1. The notation used in this supplement is the same as that in the main paper.

1 Preliminaries

Before presenting the proofs, we review the following elementary results for both multivariate normal distributions and Wishart distributions which are use extensively in the proofs. Let $X \sim N_p(\mu, \Sigma)$. Then

$$E\|X - c\|^2 = \operatorname{tr}\Sigma + \|\mu - c\|^2$$
$$E\|X - c\|^4 = (\operatorname{tr}\Sigma)^2 + 2\operatorname{tr}(\Sigma^2) + 4(\mu - c)^T\Sigma(\mu - c) + 2\|\mu - c\|^2 \operatorname{tr}\Sigma + \|\mu - c\|^4$$

where $c \in \mathbb{R}^p$. For $X \sim \mathrm{LN}(M, \Sigma)$,

$$Ed^2_{\mathrm{LE}}(X, C) = \operatorname{tr}\Sigma + d^2_{\mathrm{LE}}(M, C)$$
$$Ed^4_{\mathrm{LE}}(X, C) = (\operatorname{tr}\Sigma)^2 + 2\operatorname{tr}(\Sigma^2) + 4(\widetilde{M} - \widetilde{C})^T\Sigma(\widetilde{M} - \widetilde{C}) + 2d^2_{\mathrm{LE}}(M, C)\operatorname{tr}\Sigma + d^4_{\mathrm{LE}}(M, C)$$

where $C \in P_N$. These results can be easily obtained from the definition of the Log-Normal distribution.

Let Y be a $p \times p$ symmetric matrix with eigenvalues $\lambda_1, \dots, \lambda_p$ and let $\kappa = (k_1, \dots, k_p)$ be a (non-increasing) partition of a positive integer k, i.e. $k_1 \geq k_2 \geq, \dots, \geq k_p$ and $\sum_{i=1}^p k_i = k$ where the k_i's are non-negative integers. The *zonal polynomial* of Y corresponding to κ, denoted by $C_\kappa(Y)$, is a symmetric, homogeneous polynomial of degree k in the eigenvalues $\lambda_1, \dots, \lambda_p$. One of the properties of the zonal polynomials is that $\big(\operatorname{tr} Y\big)^k = \sum_\kappa C_\kappa(Y)$. For a more precise definition of the zonal polynomial and its calculations, we refer the readers to Ch. 7 of Muirhead (1982). The following lemma is essential in our work.

Lemma 1 *(Muirhead (1982), Corollary 7.2.4) If Y is positive definite, then $C_\kappa(Y) > 0$ for all partitions κ.*

The jth *elementary symmetric function* of Y, denoted by $\operatorname{tr}_j Y$, is the sum of all principal minors of order j of the matrix Y. For the case of $j = 1$, $\operatorname{tr}_1 Y = \sum_{i=1}^{p} \lambda_i = \operatorname{tr} Y$, and for the case of $j = 2$, $\operatorname{tr}_2 Y = \sum_{i<j} \lambda_i\lambda_j$. The definition gives rise to the following identities which are useful in the proofs

$$(\operatorname{tr} Y)^2 = \operatorname{tr}(Y^2) + 2\operatorname{tr}_2 Y \tag{1}$$

$$\begin{aligned}(\operatorname{tr} Y)^4 &= (\operatorname{tr}(Y^2))^2 + 4\operatorname{tr}(Y^2)\operatorname{tr}_2 Y + 4(\operatorname{tr}_2 Y)^2 \\ &= \operatorname{tr}(Y^4) + 2\operatorname{tr}_2 Y^2 + 2(\operatorname{tr}_2 Y)(\operatorname{tr} Y)^2 \end{aligned}\tag{2}$$

For $S \sim \text{Wishart}_p(\Sigma, \nu)$, the kth moment, $k = 0, 1, 2, \ldots$ of $\operatorname{tr} S$ is given by

$$E(\operatorname{tr} S)^k = 2^k \sum_{\kappa} \left(\frac{\nu}{2}\right)_\kappa C_\kappa(\Sigma) \tag{3}$$

where

$$(a)_\kappa = \prod_{i=1}^{p} \left(a - \frac{i-1}{2}\right)_{k_i}$$

is called the *generalized hypergeometric coefficient* and

$$(a)_{k_i} = a(a+1)\ldots(a+k_i-1),\ (a)_0 = 1$$

(Gupta & Nagar (2000), Theorem 3.3.23). Next, we review some elementary results for the Wishart distribution:

$$\begin{aligned} ES &= \nu\Sigma \\ ES^2 &= \nu(\nu+1)\Sigma^2 + \nu(\operatorname{tr}\Sigma)\Sigma \\ E\operatorname{tr}_k S &= \nu(\nu-1)\cdots(\nu-k+1)\operatorname{tr}_k \Sigma \\ E(\operatorname{tr} S)^2 &= \nu(\nu+2)(\operatorname{tr}\Sigma)^2 - 4\nu\operatorname{tr}_2\Sigma = \nu^2(\operatorname{tr}\Sigma)^2 + 2\nu\operatorname{tr}(\Sigma^2). \end{aligned}$$

These results can be found in Gupta & Nagar (2000)(p. 99 and p. 106).

2 Derivation of the Risk Function and the SURE

Recall that the loss function is $L\big((\widehat{\boldsymbol{M}}, \widehat{\boldsymbol{\Sigma}}), (\boldsymbol{M}, \boldsymbol{\Sigma})\big) = p^{-1}\sum_{i=1}^{p} d^2_{\text{LE}}(\widehat{M}_i, M_i) + p^{-1}\sum_{i=1}^{p} \|\widehat{\Sigma}_i - \Sigma_i\|^2 = L_1(\widehat{\boldsymbol{M}}, \boldsymbol{M}) + L_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$. Write $R\big((\widehat{\boldsymbol{M}}, \widehat{\boldsymbol{\Sigma}}), (\boldsymbol{M}, \boldsymbol{\Sigma})\big) = EL_1(\widehat{\boldsymbol{M}}, \boldsymbol{M}) + EL_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = R_1(\widehat{\boldsymbol{M}}, \boldsymbol{M}) + R_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma})$. Then

$$\begin{aligned} R_1(\widehat{\boldsymbol{M}}, \boldsymbol{M}) &= p^{-1}\sum_{i=1}^{p} Ed^2_{\text{LE}}(\widehat{M}_i, M_i) \\ &= p^{-1}\sum_{i=1}^{p}\left[\frac{n^2}{(\lambda+n)^2} Ed^2_{\text{LE}}(\bar{X}_i, M_i) + \frac{\lambda^2}{(\lambda+n)^2} d^2_{\text{LE}}(\mu, M_i)\right] \\ &= p^{-1}(\lambda+n)^{-2}\sum_{i=1}^{p}\left[n\operatorname{tr}\Sigma_i + \lambda^2 d^2_{\text{LE}}(\mu, M_i)\right] \end{aligned}$$

and

$$
\begin{aligned}
R_2(\widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= p^{-1} \sum_{i=1}^{p} (\nu + n - q - 2)^{-2} E \|(\Psi - (\nu - q - 1)\Sigma_i) + (S_i - (n-1)\Sigma_i)\|^2 \\
&= p^{-1} \sum_{i=1}^{p} (\nu + n - q - 2)^{-2} \Big[E \operatorname{tr}(S_i^2) - 2(n-1) E \operatorname{tr}(S_i \Sigma_i) + (n-1)^2 \operatorname{tr}(\Sigma_i^2) \\
&\quad + \operatorname{tr}(\Psi^2) - 2(\nu - q - 1) \operatorname{tr}(\Psi \Sigma_i) + (\nu - q - 1)^2 \operatorname{tr}(\Sigma_i^2) \Big] \\
&= p^{-1} \sum_{i=1}^{p} (\nu + n - q - 2)^{-2} \Big[n(n-1) \operatorname{tr}(\Sigma_i^2) + (n-1)(\operatorname{tr} \Sigma_i)^2 - 2(n-1)^2 \operatorname{tr}(\Sigma_i^2) \\
&\quad + (n-1)^2 \operatorname{tr}(\Sigma_i^2) + \operatorname{tr}(\Psi^2) - 2(\nu - q - 1) \operatorname{tr}(\Psi \Sigma_i) + (\nu - q - 1)^2 \operatorname{tr}(\Sigma_i^2) \Big] \\
&= p^{-1} \sum_{i=1}^{p} (\nu + n - q - 2)^{-2} \Big[\big(n - 1 + (\nu - q - 1)^2 \big) \operatorname{tr}(\Sigma_i^2) \\
&\quad - 2(\nu - q - 1) \operatorname{tr}(\Psi \Sigma_i) + (n-1)(\operatorname{tr}\Sigma_i)^2 + \operatorname{tr}(\Psi^2) \Big].
\end{aligned}
$$

To obtain the SURE for this risk function, we first have to find unbiased estimates of the quantities $\operatorname{tr}\Sigma_i$, $d^2_{\mathrm{LE}}(\mu, M_i)$, $\operatorname{tr}(\Sigma_i^2)$, $\operatorname{tr}(\Psi\Sigma_i)$, and $(\operatorname{tr}\Sigma_i)^2$. With the results provided in Section 1, it is easy to verify the following equations:

$$
\begin{aligned}
\operatorname{tr} \Sigma_i &= E\big((n-1)^{-1} \operatorname{tr} S_i \big) \\
\operatorname{tr}(\Psi \Sigma_i) &= E\big((n-1)^{-1} \operatorname{tr}(\Psi S_i) \big) \\
(\operatorname{tr} \Sigma_i)^2 &= E\Big(\frac{n (\operatorname{tr} S_i)^2 - 2 \operatorname{tr} S_i^2}{(n-1)(n+1)(n-2)} \Big) \\
\operatorname{tr} \Sigma_i^2 &= E\Big(\frac{(n-1) \operatorname{tr} S_i^2 - (\operatorname{tr} S_i)^2}{(n-1)(n+1)(n-2)} \Big) \\
d^2_{\mathrm{LE}}(\mu, M_i) &= E\Big(d^2_{\mathrm{LE}}(\bar{X}_i, \mu) - \frac{\operatorname{tr} S_i}{n(n-1)} \Big).
\end{aligned}
$$

Plugging the above unbiased estimates into the risk function we obtain

$$
\begin{aligned}
\mathrm{SURE}(\lambda,\Psi,\nu,\mu) = p^{-1}\sum_{i=1}^{p}\Bigg\{&(\lambda+n)^{-2}\Big[\frac{n}{n-1}\operatorname{tr}S_i+\lambda^2 d_{\mathrm{LE}}^2(\bar{X}_i,\mu)-\frac{\lambda^2}{n(n-1)}\operatorname{tr}S_i\Big] \\
&+(\nu+n-q-2)^{-2}\Big[(n-1+(\nu-q-1)^2)\Big(\frac{(n-1)\operatorname{tr}S_i^2-(\operatorname{tr}S_i)^2}{(n-1)(n+1)(n-2)}\Big) \\
&+\frac{n(\operatorname{tr}S_i)^2-2\operatorname{tr}S_i^2}{(n+1)(n-2)}-2\frac{\nu-q-1}{n-1}\operatorname{tr}(\Psi S_i)+\operatorname{tr}(\Psi^2)\Big]\Bigg\} \\
= p^{-1}\Bigg\{&\sum_{i=1}^{p}(\lambda+n)^{-2}\Big[\frac{n-\lambda^2/n}{n-1}\operatorname{tr}S_i+\lambda^2 d_{\mathrm{LE}}^2(\bar{X}_i,\mu)\Big] \\
&+(\nu+n-q-2)^{-2}\Big[\frac{n-3+(\nu-q-1)^2}{(n+1)(n-2)}\operatorname{tr}(S_i^2) \\
&+\frac{(n-1)^2-(\nu-q-1)^2}{(n-1)(n+1)(n-2)}\big(\operatorname{tr}S_i\big)^2-2\frac{\nu-q-1}{n-1}\operatorname{tr}(\Psi S_i)+\operatorname{tr}(\Psi^2)\Big]\Bigg\}.
\end{aligned}
$$

3 Proofs of the Theorems

The following lemmas are essential for proving Theorem 3.

Lemma 2 *(Xie et al. 2012) Let* $X_i \overset{ind}{\sim} N(\theta_i, A_i)$*. Assume the following conditions:*

(i) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p} A_i^2 < \infty$,

(ii) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p} A_i\theta_i^2 < \infty$,

(iii) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p} |\theta_i|^{2+\delta} < \infty$ *for some* $\delta > 0$.

Then $E\big(\max_{1\le i\le p} X_i^2\big) = O(p^{2/(2+\delta^*)})$ *where* $\delta^* = \min(1,\delta)$.

The next lemma is an extension of the previous lemma to Log-Normal distributions.

Lemma 3 *Let* $X_i \overset{ind}{\sim} LN(M_i, \Sigma_i)$ *on* P_N *and* $q = N(N+1)/2$*. Assume the following conditions:*

(i) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p} \big(\operatorname{tr}\Sigma_i\big)^2 < \infty$,

(ii) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p} \widetilde{M}_i^T\Sigma_i\widetilde{M}_i < \infty$,

(iii) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p} \|\log M_i\|^{2+\delta} < \infty$ *for some* $\delta > 0$.

Then $E\big(\max_{1\le i\le p} \|\log X_i\|^2\big) = O(p^{2/(2+\delta^*)})$ *where* $\delta^* = \min(1,\delta)$.

Proof. Write $Y_i = \widetilde{X}_i$ and $\mu_i = \widetilde{M}_i$. Then $\|Y_i\|^2 = \|\log X_i\|^2$. From the definition of the Log-Normal distribution, $Y_i \overset{\text{ind}}{\sim} N_q(\mu_i, \Sigma_i)$. Since for $j = 1, \dots, q$

$$\sum_{i=1}^{p} \Sigma_{i,jj}^2 < \sum_{i=1}^{p} \big(\operatorname{tr}\Sigma_i\big)^2$$
$$\sum_{i=1}^{p} \Sigma_{i,jj}\mu_{i,j}^2 < \sum_{i=1}^{p} \mu_i^T \Sigma_i \mu_i = \sum_{i=1}^{p} \widetilde{M}_i^T \Sigma_i \widetilde{M}_i$$
$$\sum_{i=1}^{p} |\mu_{i,j}|^{2+\delta} < \sum_{i=1}^{p} \|\mu_i\|^{2+\delta} = \sum_{i=1}^{p} \|\log M_i\|^{2+\delta},$$

by Lemma 2, we have $E(\max_{1\le i \le, p} Y_{i,j}^2) = O(p^{2/(2+\delta^*)})$. Then

$$E\Big(\max_{1\le i\le p} \|\log X_i\|^2\Big) = E\Big(\max_{1\le i\le p} \|Y_i\|^2\Big)$$
$$\le E\Big(\sum_{j=1}^{q} \max_{1\le i\le p} Y_{i,j}^2\Big) = \sum_{j=1}^{q} E\Big(\max_{1\le i\le p} Y_{i,j}^2\Big) = O(p^{2/(2+\delta *)})$$

which concludes the proof. □

Lemma 4 *Let* $S_i \overset{ind}{\sim}$ *Wishart*(Σ_i, ν) *where the* Σ_i*'s are* $q \times q$ *symmetric positive-definite matrices. If* $\limsup_{p\to\infty} p^{-1} \sum_{i=1}^{p} (\operatorname{tr}\Sigma_i)^4 < \infty$, *then* $E(\max_{1\le i\le p} \|S_i\|^2) = O(q^2 p^{1/2} (\log p)^2 + q^2 p^{1/2} (\log q)^2)$.

Proof. Write $S_i = X_i X_i^T$ where $X_i \overset{\text{ind}}{\sim} N_q(0, \Sigma_i)$. Then $\|S_i\|^2 = \|X_i\|^4$. From Lemma 2, we have $E(\max_{1\le j\le q} X_{i,j}^2) = O(p^{2/3})$ and $E(\max_{1\le i\le p, 1\le j\le q} X_{i,j}^2) = O(q^{2/3} p^{2/3})$. Let $X_{i,j} = \Sigma_{i,j}^{1/2} Z_{i,j}$ where $Z_{i,j} \overset{\text{iid}}{\sim} N(0,1)$. Then $X_{i,j}^4 = \Sigma_{i,jj}^2 Z_{i,j}$ and

$$\max_{1\le i\le p, 1\le j\le q} X_{i,j}^4 \le \max_{1\le i\le p, q\le j\le q} \Sigma_{i,jj}^2 \cdot \max_{1\le i\le p, 1\le j\le q} Z_{i,j}.$$

Since

$$\max_{1\le i\le p, 1\le j\le q} \Sigma_{i,jj}^4 < \max_{1\le i\le p} \operatorname{tr}(\Sigma_i)^4 < \sum_{i=1}^{p} \operatorname{tr}(\Sigma_i)^4 = O(p)$$

implies $\max_{1\le i\le p, 1\le j\le q} \Sigma_{i,jj}^2 = O(p^{1/2})$ and

$$E\Big(\max_{1\le i\le p, 1\le j\le q} Z_{i,j}^4\Big) = O((\log p + \log q)^2),$$

we have

$$E\Big(\max_{1\le i\le p, 1\le j\le q} X_{i,j}^4\Big) \le \max_{1\le i\le p, q\le j\le q} \Sigma_{i,jj}^2 E\Big(\max_{1\le i\le p, 1\le j\le q} Z_{i,j}\Big) = O(p^{1/2}(\log p + \log q)^2).$$

Then

$$
\begin{aligned}
E\Big(\max_{1\le i\le p}\|S_i\|^2\Big) &= E\Big(\max_{1\le i\le p}\|X_i\|^4\Big) = E\left[\max_{1\le i\le p}\left(\sum_{j=1}^{q} X_{i,j}^2\right)^2\right] \\
&= E\left[\max_{1\le i\le p}\left(\sum_{j=1}^{q} X_{i,j}^4 + \sum_{j\neq k} X_{i,j}^2 X_{i,k}^2\right)\right] \\
&\le qE\Big(\max_{1\le i\le p, 1\le j\le q} X_{i,j}^4\Big) + q(q-1)E\Big(\max_{1\le i\le p, 1\le j\le q} X_{i,j}^4\Big) \\
&\le q^2 O(p^{1/2}(\log p + \log q)^2) \\
&= O(q^2 p^{1/2}(\log p)^2 + q^2 p^{1/2}(\log q)^2)
\end{aligned}
$$

□

Theorem 3 *Assume the following conditions:*

(i) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p}\big(\operatorname{tr}\Sigma_i\big)^4 < \infty$,

(ii) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p}\widetilde{M}_i^T\Sigma_i\widetilde{M}_i < \infty$,

(iii) $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^{p}\|\log M_i\|^{2+\delta} < \infty$ *for some* $\delta > 0$.

Then

$$
\sup_{\substack{\lambda>0,\nu>q+1,\|\Psi\|\le\max_{1\le i\le p}\|S_i\|,\\ \|\log\mu\|\le\max_{1\le i\le p}\|\log \bar{X}_i\|}} \Big| SURE(\lambda,\Psi,\nu,\mu) - L\Big(\big(\widehat{\boldsymbol{M}}^{\lambda,\mu}, \widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}\big), \big(\boldsymbol{M},\boldsymbol{\Sigma}\big)\Big)\Big| \xrightarrow{prob} 0 \quad \text{as } p\to\infty.
$$

Proof. First, we write the loss function L as

$$
\begin{aligned}
L\Big(\big(\widehat{\boldsymbol{M}}^{\lambda,\mu}, \widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}\big), \big(\boldsymbol{M},\boldsymbol{\Sigma}\big)\Big) &= p^{-1}\sum_{i=1}^{p} d_{\mathrm{LE}}^2(\widehat{M}_i^{\lambda,\mu}, M_i) + \|\widehat{\Sigma}_i^{\Psi,\nu} - \Sigma_i\|^2 \\
&= L_1\Big(\widehat{\boldsymbol{M}}^{\lambda,\mu}, \boldsymbol{M}\Big) + L_2\Big(\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}, \boldsymbol{\Sigma}\Big),
\end{aligned}
$$

where

$$
\begin{aligned}
L_1\Big(\widehat{\boldsymbol{M}}^{\lambda,\mu},\boldsymbol{M}\Big) &= p^{-1}\sum_{i=1}^{p}\Big\|(\lambda+n)^{-1}\Big(n\Big(\log\bar{X}_i-\log M_i\Big)+\lambda\Big(\log\mu-\log M_i\Big)\Big)\Big\|^2\\
&= p^{-1}\sum_{i=1}^{p}(\lambda+n)^{-2}\Big[n^2 d_{\mathrm{LE}}^2(\bar{X}_i,M_i)+\lambda^2 d_{\mathrm{LE}}^2(\mu,M_i)\\
&\quad+2n\lambda\Big\langle\log\bar{X}_i-\log M_i,\log\mu-\log M_i\Big\rangle\Big],\\
L_2\Big(\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu},\boldsymbol{\Sigma}\Big) &= p^{-1}\sum_{i=1}^{p}(\nu+n-q-2)^{-2}\Big\|\big(\Psi-(\nu-q-1)\Sigma_i\big)+\big(S_i-(n-1)\Sigma_i\big)\Big\|^2\\
&= p^{-1}\sum_{i=1}^{p}(\nu+n-q-2)^{-2}\Big[\operatorname{tr}\big(\Psi^2\big)-2(\nu-q-1)\operatorname{tr}\big(\Psi\Sigma_i\big)+(\nu-q-1)^2\operatorname{tr}\big(\Sigma_i^2\big)\\
&\quad+\operatorname{tr}\big(S_i^2\big)-2(n-1)\operatorname{tr}\big(S_i\Sigma_i\big)+(n-1)^2\operatorname{tr}\big(\Sigma_i^2\big)\\
&\quad+2\langle\Psi-(\nu-q-1)\Sigma_i,S_i-(n-1)\Sigma_i\rangle\Big].
\end{aligned}
$$

Write the SURE as

$$
\mathrm{SURE}(\lambda,\mu,\Psi,\nu)=\mathrm{SURE}_1(\lambda,\mu)+\mathrm{SURE}_2(\Psi,\nu),
$$

where

$$
\begin{aligned}
\mathrm{SURE}_1(\lambda,\mu) &= p^{-1}\sum_{i=1}^{p}(\lambda+n)^{-2}\Big[\frac{n-\lambda^2/n}{n-1}\operatorname{tr}S_i+\lambda^2 d_{\mathrm{LE}}^2(\bar{X}_i,\mu)\Big],\\
\mathrm{SURE}_2(\Psi,\nu) &= p^{-1}\sum_{i=1}^{p}(\nu+n-q-2)^{-2}\Bigg[\frac{n-3+(\nu-q-1)^2}{(n+1)(n-2)}\operatorname{tr}\big(S_i^2\big)\\
&\quad+\frac{(n-1)^2-(\nu-q-1)^2}{(n-2)(n-1)(n+1)}\big(\operatorname{tr}S_i\big)^2-2\frac{\nu-q-1}{n-1}\operatorname{tr}\big(\Psi S_i\big)+\operatorname{tr}\big(\Psi^2\big)\Bigg].
\end{aligned}
$$

Since

$$
\begin{aligned}
&\sup_{\substack{\lambda>0,\nu>q+1,\|\Psi\|\le\max_{1\le i\le p}\|S_i\|,\\ \|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}}\Big|\mathrm{SURE}(\lambda,\Psi,\nu,\mu)-L\Big(\big(\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}\big),\big(\boldsymbol{M},\boldsymbol{\Sigma}\big)\Big)\Big|\le\\
&\qquad\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}\Big|\mathrm{SURE}_1(\lambda,\mu)-L_1\big(\widehat{\boldsymbol{M}}^{\lambda,\mu},\boldsymbol{M}\big)\Big|\\
&\qquad\qquad+\sup_{\nu>q+1,\|\Psi\|\le\max_{1\le i\le p}\|S_i\|}\Big|\mathrm{SURE}_2(\Psi,\nu)-L_2\big(\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu},\boldsymbol{\Sigma}\big)\Big|,
\end{aligned}
$$

it suffices to show the two terms on the right-hand side converge to 0 in probability. For the first term,

$$\begin{aligned}
|\mathrm{SURE}_1(\lambda,\mu) - L_1(\widehat{\boldsymbol{M}}^{\lambda,\mu}, \boldsymbol{M})| &= \Bigg| p^{-1}\sum_{i=1}^{p}(\lambda+n)^{-2}\Bigg[\frac{n}{n-1}\operatorname{tr} S_i - n^2 d_{\mathrm{LE}}^2(\bar{X}_i, M_i) \\
&\quad + \lambda^2\Big(\frac{\operatorname{tr} S_i}{n(n-1)} + d_{\mathrm{LE}}^2(\bar{X}_i,\mu) - d_{\mathrm{LE}}^2(M_i,\mu)\Big) \\
&\quad + 2n\lambda\Big\langle \log\bar{X}_i - \log M_i, \log\mu - \log M_i\Big\rangle\Bigg]\Bigg| \\
&\le \Bigg| p^{-1}\sum_{i=1}^{p}(\lambda+n)^{-2}\Big(\frac{n}{n-1}\operatorname{tr} S_i - n^2 d_{\mathrm{LE}}^2(\bar{X}_i, M_i)\Big)\Bigg| \qquad (4) \\
&\quad + \Bigg| p^{-1}\sum_{i=1}^{p}\frac{\lambda^2}{(\lambda+n)^2}\Big(\frac{\operatorname{tr} S_i}{n(n-1)} + d_{\mathrm{LE}}^2(\bar{X}_i,\mu) - d_{\mathrm{LE}}^2(M_i,\mu)\Big)\Bigg| \qquad (5) \\
&\quad + \Bigg| p^{-1}\sum_{i=1}^{p}\frac{2n\lambda}{(\lambda+n)^2}\Big\langle \log\bar{X}_i - \log M_i, \log\mu - \log M_i\Big\rangle\Bigg|. \qquad (6)
\end{aligned}$$

We will now prove the convergence of each of the three terms individually.

For (4), from assumption (i), we have

$$\begin{aligned}
&\operatorname{Var}\left(p^{-1}\sum_{i=1}^{p}\Big(\frac{n}{n-1}\operatorname{tr} S_i - n^2 d_{\mathrm{LE}}^2(\bar{X}_i, M_i)\Big)\right) \\
&= \frac{1}{p}p^{-1}\sum_{i=1}^{p}\operatorname{Var}\Big(\frac{n}{n-1}\operatorname{tr} S_i - n^2 d_{\mathrm{LE}}^2(\bar{X}_i, M_i)\Big) \\
&= \frac{1}{p}p^{-1}\sum_{i=1}^{p}E\Big(\frac{n}{n-1}\operatorname{tr} S_i - n^2 d_{\mathrm{LE}}^2(\bar{X}_i, M_i)\Big)^2 \\
&= \frac{n^2}{p}p^{-1}\sum_{i=1}^{p}\Big[\frac{E\big(\operatorname{tr} S_i\big)^2}{(n-1)^2} + n^2 E d_{\mathrm{LE}}^4(\bar{X}_i, M_i) - 2\frac{n}{n-1}E\big(\operatorname{tr} S_i\big)E d_{\mathrm{LE}}^2(\bar{X}_i, M_i)\Big] \\
&= \frac{n^2}{p}p^{-1}\sum_{i=1}^{p}\Big[\frac{n+1}{n-1}\big(\operatorname{tr}\Sigma_i\big)^2 + 4\frac{\operatorname{tr}_2\Sigma_i}{n-1} + \big(\operatorname{tr}\Sigma_i\big)^2 + 2\operatorname{tr}\big(\Sigma_i^2\big) - 2\big(\operatorname{tr}\Sigma_i\big)^2\Big] \\
&= \frac{n^2}{p}p^{-1}\sum_{i=1}^{p}\Big[\frac{2}{n-1}\big(\operatorname{tr}\Sigma_i\big)^2 + \frac{4}{n-1}\operatorname{tr}_2\Sigma_i + 2\operatorname{tr}\big(\Sigma_i^2\big)\Big] \overset{p\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Then, by Markov's inequality,

$$\left|p^{-1}\sum_{i=1}^{p}\Big(\frac{n}{n-1}\operatorname{tr} S_i - n^2 d_{\mathrm{LE}}^2(\bar{X}_i, M_i)\Big)\right| \overset{\text{prob}}{\longrightarrow} 0 \quad \text{as } p\to\infty.$$

Thus

$$
\begin{aligned}
&\sup_{\lambda>0}\left|p^{-1}\sum_{i=1}^{p}(\lambda+n)^{-2}\Big(\frac{n}{n-1}\operatorname{tr}S_i-n^2d_{\mathrm{LE}}^2(\bar{X}_i,M_i)\Big)\right| \\
&=\Big(\sup_{\lambda>0}(\lambda+n)^{-2}\Big)\left|p^{-1}\sum_{i=1}^{p}\Big(\frac{n}{n-1}\operatorname{tr}S_i-n^2d_{\mathrm{LE}}^2(\bar{X}_i,M_i)\Big)\right| \\
&=\frac{1}{n^2}\left|p^{-1}\sum_{i=1}^{p}\Big(\frac{n}{n-1}\operatorname{tr}S_i-n^2d_{\mathrm{LE}}^2(\bar{X}_i,M_i)\Big)\right|\xrightarrow{\text{prob}}0\quad\text{as } p\to\infty. \qquad (7)
\end{aligned}
$$

Remark By the identity (1), assumption (i) implies $\limsup_{p\to\infty}p^{-1}\sum_{i=1}^{p}\operatorname{tr}\left(\Sigma_i^2\right)<\infty$ and $\limsup_{p\to\infty}p^{-1}\sum_{i=1}^{p}\operatorname{tr}_2\Sigma_i<\infty$.

For (5),

$$
\begin{aligned}
&\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}\left|p^{-1}\sum_{i=1}^{p}\frac{\lambda^2}{(\lambda+n)^2}\Big(\frac{\operatorname{tr}S_i}{n(n-1)}+d_{\mathrm{LE}}^2(\bar{X}_i,\mu)-d_{\mathrm{LE}}^2(M_i,\mu)\Big)\right| \\
&=\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}\Bigg|p^{-1}\sum_{i=1}^{p}\frac{\lambda^2}{(\lambda+n)^2}\Big(\frac{\operatorname{tr}S_i}{n(n-1)}+\|\log\bar{X}_i\|^2-\|\log M_i\|^2 \\
&\qquad+2\langle\log\bar{X}_i-\log M_i,\log\mu\rangle\Big)\Bigg| \\
&\le\sup_{\lambda>0}\left|p^{-1}\sum_{i=1}^{p}\frac{\lambda^2}{(\lambda+n)^2}\Big(\frac{\operatorname{tr}S_i}{n(n-1)}+\|\log\bar{X}_i\|^2-\|\log M_i\|^2\Big)\right| \\
&\qquad+\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}\left|p^{-1}\frac{2\lambda^2}{(\lambda+n)^2}\Big\langle\sum_{i=1}^{p}\log\bar{X}_i-\log M_i,\log\mu\Big\rangle\right| \\
&\le\Big(\sup_{\lambda>0}\frac{\lambda^2}{(\lambda+n)^2}\Big)\left|p^{-1}\sum_{i=1}^{p}\Big(\frac{\operatorname{tr}S_i}{n(n-1)}+\|\log\bar{X}_i\|^2-\|\log M_i\|^2\Big)\right| \\
&\qquad+\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}\left|p^{-1}\frac{2\lambda^2}{(\lambda+n)^2}\|\log\mu\|\left\|\sum_{i=1}^{p}\big(\log\bar{X}_i-\log M_i\big)\right\|\right| \\
&\hspace{20em}\text{(By Cauchy's inequality)} \\
&\le\left|p^{-1}\sum_{i=1}^{p}\Big(\frac{\operatorname{tr}S_i}{n(n-1)}+\|\log\bar{X}_i\|^2-\|\log M_i\|^2\Big)\right| \\
&\qquad+\left|p^{-1}\max_{1\le i\le p}\|\log\bar{X}_i\|\left\|\sum_{i=1}^{p}\big(\log\bar{X}_i-\log M_i\big)\right\|\right|
\end{aligned}
$$

Since by assumptions (i) and (ii),

$$
\begin{aligned}
&\mathrm{Var}\left(p^{-1}\sum_{i=1}^{p}\Big(\frac{\operatorname{tr} S_i}{n(n-1)}+\|\log \bar{X}_i\|^2-\|\log M_i\|^2\Big)\right)\\
&= p^{-2}\sum_{i=1}^{p}\mathrm{Var}\Big(\frac{\operatorname{tr} S_i}{n(n-1)}+\|\log \bar{X}_i\|^2-\|\log M_i\|^2\Big)\\
&= p^{-2}\sum_{i=1}^{p}E\Big(\frac{\operatorname{tr} S_i}{n(n-1)}+\|\log \bar{X}_i\|^2-\|\log M_i\|^2\Big)^2\\
&= p^{-2}\sum_{i=1}^{p}\Bigg[\left(\frac{n+1}{n^2(n-1)}\big(\operatorname{tr}\Sigma_i\big)^2-\frac{4}{n^2(n-1)}\operatorname{tr}_2\Sigma_i\right)\left(+\frac{\big(\operatorname{tr}\Sigma_i\big)^2}{n^2}+\frac{2\operatorname{tr}\big(\Sigma_i^2\big)}{n^2}\right.\\
&\qquad\left.+\frac{4}{n}\widetilde{M}_i^T\Sigma_i\widetilde{M}_i+\frac{2}{n}\|\log M_i\|^2\operatorname{tr}\Sigma_i+\|\log M_i\|^4\right)+\|\log M_i\|^4\\
&\qquad+\frac{2}{n}\operatorname{tr}\Sigma_i\Big(\frac{1}{n}\operatorname{tr}\Sigma_i+\|\log M_i\|^2\Big)-2\Big(\frac{1}{n}\operatorname{tr}\Sigma_i+\|\log M_i\|^2\Big)\|\log M_i\|^2-\frac{2}{n}\operatorname{tr}\Sigma_i\|\log M_i\|^2\Bigg]\\
&= p^{-2}\sum_{i=1}^{p}\left[\frac{4n-2}{n^2(n-1)}\big(\operatorname{tr}\Sigma_i\big)^2-\frac{4}{n^2(n-1)}\operatorname{tr}_2\Sigma_i+\frac{2}{n}\operatorname{tr}\big(\Sigma_i^2\big)+\frac{4}{n}\widetilde{M}_i^T\Sigma_i\widetilde{M}_i\right]\xrightarrow{p\to\infty}0,
\end{aligned}
$$

we have

$$
\left|p^{-1}\sum_{i=1}^{p}\Big(\frac{\operatorname{tr} S_i}{n(n-1)}+\|\log \bar{X}_i\|^2-\|\log M_i\|^2\Big)\right|\xrightarrow{\text{prob}}0\quad\text{as } p\to\infty
$$

by Markov's inequality. Since by Lemma 3,

$$
\begin{aligned}
&E\left[\frac{2}{p}\max_{1\le i\le p}\|\log \bar{X}_i\|\left\|\sum_{i=1}^{p}\big(\log \bar{X}_i-\log M_i\big)\right\|\right]\\
&\le \frac{2}{p}\left[E\Big(\max_{1\le i\le p}\|\log \bar{X}_i\|^2\Big)E\left\|\sum_{i=1}^{p}\big(\log \bar{X}_i-\log M_i\big)\right\|^2\right]^{1/2}\\
&= O(p^{-1})\times O(p^{1/(2+\delta^*)})\times O(p^{1/2})\\
&= O(p^{-\delta^*/(4+2\delta^*)}),
\end{aligned}
$$

we have

$$
\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log \bar{X}_i\|}\left|p^{-1}\sum_{i=1}^{p}\frac{\lambda^2}{(\lambda+n)^2}\Big(\frac{\operatorname{tr} S_i}{n(n-1)}+d_{\mathrm{LE}}^2(\bar{X}_i,\mu)-d_{\mathrm{LE}}^2(M_i,\mu)\Big)\right|\xrightarrow{\text{prob}}0 \quad \text{as } p\to\infty. \tag{8}
$$

For (6), we have

$$
\begin{aligned}
&\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}\left|p^{-1}\sum_{i=1}^{p}\frac{2n\lambda}{(\lambda+n)^2}\Big\langle\log\bar{X}_i-\log M_i,\log\mu-\log M_i\Big\rangle\right|\\
&\le\sup_{\lambda>0}\left|p^{-1}\frac{2n\lambda}{(\lambda+n)^2}\max_{1\le i\le p}\|\log\bar{X}_i\|\left\|\sum_{i=1}^{p}\big(\log\bar{X}_i-\log M_i\big)\right\|\right|\\
&\quad+\sup_{\lambda>0}\left|p^{-1}\sum_{i=1}^{p}\frac{2n\lambda}{(\lambda+n)^2}\Big\langle\log\bar{X}_i-\log M_i,\log M_i\Big\rangle\right|\\
&=\left|\frac{1}{2p}\max_{1\le i\le p}\|\log\bar{X}_i\|\left\|\sum_{i=1}^{p}\big(\log\bar{X}_i-\log M_i\big)\right\|\right|+\left|\frac{1}{2p}\sum_{i=1}^{p}\Big\langle\log\bar{X}_i-\log M_i,\log M_i\Big\rangle\right|
\end{aligned}
$$

since $\sup_{\lambda>0}2n\lambda/(\lambda+n)^2=1/2$. By assumption (ii), we have

$$
\begin{aligned}
&\operatorname{Var}\left[p^{-1}\sum_{i=1}^{p}\Big\langle\log\bar{X}_i-\log M_i,\log M_i\Big\rangle\right]\\
&=p^{-2}\sum_{i=1}^{p}E\Big\langle\log\bar{X}_i-\log M_i,\log M_i\Big\rangle^2\\
&=p^{-2}\sum_{i=1}^{p}E\widetilde{M}_i^T\left[\Big(\widetilde{\bar{X}}_i-\widetilde{M}_i\Big)\Big(\widetilde{\bar{X}}_i-\widetilde{M}_i\Big)^T\right]\widetilde{M}_i\\
&=p^{-2}\sum_{i=1}^{p}\frac{1}{n}\widetilde{M}_i^T\Sigma_i\widetilde{M}_i\overset{p\to\infty}{\longrightarrow}0
\end{aligned}
$$

and again by Markov's inequality,

$$
\left|\frac{1}{2p}\sum_{i=1}^{p}\Big\langle\log\bar{X}_i-\log M_i,\log M_i\Big\rangle\right|\overset{\text{prob}}{\longrightarrow}0\quad\text{as }p\to\infty.
$$

Thus,

$$
\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}\left|p^{-1}\sum_{i=1}^{p}\frac{2n\lambda}{(\lambda+n)^2}\Big\langle\log\bar{X}_i-\log M_i,\log\mu-\log M_i\Big\rangle\right|\overset{\text{prob}}{\longrightarrow}0\quad\text{as }p\to\infty.
\tag{9}
$$

Combining (7), (8), and (9), we have

$$
\sup_{\lambda>0,\|\log\mu\|\le\max_{1\le i\le p}\|\log\bar{X}_i\|}\left|\mathrm{SURE}_1(\lambda,\mu)-L_1\Big(\widehat{\boldsymbol{M}}^{\lambda,\mu},\boldsymbol{M}\Big)\right|\overset{\text{prob}}{\longrightarrow}0\quad\text{as }p\to\infty.
\tag{10}
$$

For the second term, we have

$$
\begin{aligned}
&|\mathrm{SURE}_2(\Psi,\nu) - L_2(\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu},\boldsymbol{\Sigma})| \\
&= \Bigg| p^{-1}\sum_{i=1}^{p}(\nu+n-q-2)^{-2}\Bigg[2(\nu-q-1)\Big(\frac{\mathrm{tr}(\Psi S_i)}{n-1}-\mathrm{tr}(\Psi\Sigma_i)\Big) \\
&\quad + \frac{(n-1)^2-(\nu-q-1)^2}{(n+1)(n-2)(n-1)}(\mathrm{tr}\, S_i)^2 - \frac{(n-1)^2-(\nu-q-1)^2}{(n+1)(n-2)}\,\mathrm{tr}(S_i^2) \\
&\quad + 2(n-1)\,\mathrm{tr}(S_i\Sigma_i) - \big((n-1)^2+(\nu-q-1)^2\big)\,\mathrm{tr}(\Sigma_i^2) \\
&\quad - 2\big\langle \Psi-(\nu-q-1)\Sigma_i, S_i-(n-1)\Sigma_i\big\rangle\Bigg]\Bigg| \\
&\le \Bigg| p^{-1}\sum_{i=1}^{p}\frac{2(\nu-q-1)(n-1)}{(\nu+n-q-2)^2}\langle\Psi, S_i-(n-1)\,\mathrm{tr}\,\Sigma_i\rangle\Bigg| &(11)\\
&\quad + \Bigg| p^{-1}\sum_{i=1}^{p}C(\nu)\Big[(\mathrm{tr}\, S_i)^2-(n-1)^2(\mathrm{tr}\,\Sigma_i)^2-2(n-1)\,\mathrm{tr}(\Sigma_i^2)\Big]\Bigg| &(12)\\
&\quad + \Bigg| p^{-1}\sum_{i=1}^{p}(n-1)C(\nu)\Big[\mathrm{tr}(S_i^2)-(n-1)(\mathrm{tr}\,\Sigma_i)^2-n(n-1)\,\mathrm{tr}(\Sigma_i^2)\Big]\Bigg| &(13)\\
&\quad + \Bigg| p^{-1}\sum_{i=1}^{p}2(n-1)\Big[\mathrm{tr}(S_i\Sigma_i)-(n-1)\,\mathrm{tr}(\Sigma_i^2)\Big]\Bigg| &(14)\\
&\quad + \Bigg| p^{-1}\sum_{i=1}^{p}\frac{2\big\langle\Psi-(\nu-q-1)\Sigma_i, S_i-(n-1)\Sigma_i\big\rangle}{(\nu+n-q-2)^2}\Bigg|. &(15)
\end{aligned}
$$

where

$$C(\nu) = (\nu+n-q-2)^{-2}\frac{(n-1)^2-(\nu-q-1)^2}{(n+1)(n-2)(n-1)}.$$

Note that $\sup_{\nu>q-1} C(\nu) = [(n+1)(n-2)(n-1)]^{-1}$.

For (11), by Lemma 4, we have

$$
\begin{aligned}
&\sup_{\nu>q+1,\|\Psi\|\le\max_{1\le i\le p}\|S_i\|}\Bigg| p^{-1}\sum_{i=1}^{p}\frac{2(\nu-q-1)(n-1)}{(\nu+n-q-2)^2}\big\langle\Psi, S_i-(n-1)\Sigma_i\big\rangle\Bigg| \\
&\le \sup_{\nu>q+1}\left(\frac{2(\nu-q-1)(n-1)}{(\nu+n-q-2)^2}p^{-1}\max_{1\le i\le p}\|S_i\|\Bigg\|\sum_{i=1}^{p}S_i-(n-1)\Sigma_i\Bigg\|\right) \\
&= \frac{2(n-1)^2}{(n-2)^2}\left(p^{-1}\max_{1\le i\le p}\|S_i\|\Bigg\|\sum_{i=1}^{p}S_i-(n-1)\Sigma_i\Bigg\|\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left[\frac{1}{p}\max_{1\le i\le p}\|S_i\|\left\|\sum_{i=1}^p S_i-(n-1)\Sigma_i\right\|\right]\\
&\le \frac{1}{p}\left[E\Big(\max_{i\le i\le p}\|S_i\|^2\Big)E\left\|\sum_{i=1}^p S_i-(n-1)\Sigma_i\right\|^2\right]^{1/2}\\
&=O(p^{-1})\times O(qp^{1/4}\log p+qp^{1/4}\log q)\times O(p^{1/2})\\
&=O(p^{-1/4}\log p)=o(1).
\end{aligned}
$$

Hence, we have

$$
\sup_{\nu>q+1,\|\Psi\|\le\max_{1\le i\le p}\|S_i\|}\left|p^{-1}\sum_{i=1}^p\frac{2(\nu-q-1)(n-1)}{(\nu+n-q-2)^2}\big\langle\Psi,S_i-(n-1)\Sigma_i\big\rangle\right|\xrightarrow{\text{prob}}0\quad\text{as }p\to\infty. \tag{16}
$$

For (12), we have

$$
\begin{aligned}
&\sup_{\nu>q+1}\left|p^{-1}\sum_{i=1}^p C(\nu)\Big[(\operatorname{tr}S_i)^2-(n-1)^2(\operatorname{tr}\Sigma_i)^2-2(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right|\\
&=\frac{1}{(n+1)(n-2)(n-1)}\left|p^{-1}\sum_{i=1}^p\Big[(\operatorname{tr}S_i)^2-(n-1)^2(\operatorname{tr}\Sigma_i)^2-2(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right|
\end{aligned}
$$

and

$$
\begin{aligned}
&\operatorname{Var}\left[p^{-1}\sum_{i=1}^p\Big[(\operatorname{tr}S_i)^2-(n-1)^2(\operatorname{tr}\Sigma_i)^2-2(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right]\\
&=p^{-2}\sum_{i=1}^p\operatorname{Var}\Big((\operatorname{tr}S_i)^2-(n-1)^2(\operatorname{tr}\Sigma_i)^2-2(n-1)\operatorname{tr}(\Sigma_i^2)\Big)\\
&=p^{-2}\sum_{i=1}^p E\Big((\operatorname{tr}S_i)^2-(n-1)^2(\operatorname{tr}\Sigma_i)^2-2(n-1)\operatorname{tr}(\Sigma_i^2)\Big)^2\\
&=p^{-2}\sum_{i=1}^p\Big[E(\operatorname{tr}S_i)^4-\Big((n-1)^2(\operatorname{tr}\Sigma_i)^2+2(n-1)\operatorname{tr}(\Sigma_i^2)\Big)^2\Big]\\
&=p^{-2}\sum_{i=1}^p\Big[2^4\sum_\kappa\Big(\frac{n-1}{2}\Big)_\kappa C_\kappa(\Sigma_i)-(n-1)^4(\operatorname{tr}\Sigma_i)^4\\
&\quad-4(n-1)^3(\operatorname{tr}\Sigma_i)^2\operatorname{tr}(\Sigma_i^2)-4(n-1)^2\big(\operatorname{tr}(\Sigma_i^2)\big)^2\Big]\xrightarrow{p\to\infty}0\quad(\text{by (2)}).
\end{aligned}
$$

By Markov's inequality, we have

$$
\left|p^{-1}\sum_{i=1}^p\Big[(\operatorname{tr}S_i)^2-(n-1)^2(\operatorname{tr}\Sigma_i)^2-2(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right|\xrightarrow{\text{prob}}0\quad\text{as }p\to\infty. \tag{17}
$$

Remark By Lemma 1, assumption (i) implies $\limsup_{p\to\infty} p^{-1}\sum_{i=1}^p C_\kappa(\Sigma_i) < \infty$ for all partitions $\kappa = (k_1, \ldots, k_q)$ with $\sum_{j=1}^q k_j \le 4$.

For (13), we have

$$\begin{aligned}
&\sup_{\nu>q+1}\left| p^{-1}\sum_{i=1}^p (n-1)C(\nu)\Big[\operatorname{tr}(S_i^2) - (n-1)(\operatorname{tr}\Sigma_i)^2 - n(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right| \\
&= \frac{1}{(n+1)(n-2)}\left| p^{-1}\sum_{i=1}^p \Big[\operatorname{tr}(S_i^2) - (n-1)(\operatorname{tr}\Sigma_i)^2 - n(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right|
\end{aligned}$$

and

$$\begin{aligned}
&\operatorname{Var}\left[p^{-1}\sum_{i=1}^p \Big[\operatorname{tr}(S_i^2) - (n-1)(\operatorname{tr}\Sigma_i)^2 - n(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right] \\
&= p^{-2}\sum_{i=1}^p \operatorname{Var}\Big(\operatorname{tr}(S_i^2) - (n-1)(\operatorname{tr}\Sigma_i)^2 - n(n-1)\operatorname{tr}(\Sigma_i^2)\Big) \\
&= p^{-2}\sum_{i=1}^p E\Big(\operatorname{tr}(S_i^2) - (n-1)(\operatorname{tr}\Sigma_i)^2 - n(n-1)\operatorname{tr}(\Sigma_i^2)\Big)^2 \\
&= p^{-2}\sum_{i=1}^p \Big[E(\operatorname{tr}(S_i^2))^2 - \Big((n-1)(\operatorname{tr}\Sigma_i)^2 + n(n-1)\operatorname{tr}(\Sigma_i^2)\Big)^2\Big] \\
&\le p^{-2}\sum_{i=1}^p \Big[E(\operatorname{tr} S_i)^4 - (n-1)^2(\operatorname{tr}\Sigma_i)^4 \\
&\quad - 2n(n-1)^2(\operatorname{tr}\Sigma_i)^2\operatorname{tr}(\Sigma_i^2) - n^2(n-1)^2\big(\operatorname{tr}(\Sigma_i^2)\big)^2\Big] \xrightarrow{p\to\infty} 0 \quad \text{(by (2))}.
\end{aligned}$$

By Markov's inequality, we have

$$\left| p^{-1}\sum_{i=1}^p \Big[\operatorname{tr}(S_i^2) - (n-1)(\operatorname{tr}\Sigma_i)^2 - n(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right| \xrightarrow{\text{prob}} 0 \quad \text{as } p\to\infty. \tag{18}$$

For (14), by Cauchy's inequality, we have

$$\begin{aligned}
\left| p^{-1}\sum_{i=1}^p 2(n-1)\Big[\operatorname{tr}(S_i\Sigma_i) - (n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right| &= \left| p^{-1}\sum_{i=1}^p 2(n-1)\langle \Sigma_i, S_i - (n-1)\Sigma_i\rangle\right| \\
&\le 2(n-1)p^{-1}\sum_{i=1}^p |\langle \Sigma_i, S_i - (n-1)\Sigma_i\rangle| \\
&\le 2(n-1)p^{-1}\sum_{i=1}^p \|\Sigma_i\|\|S_i - (n-1)\Sigma_i\|.
\end{aligned}$$

Then

$$
\begin{aligned}
&\operatorname{Var}\left(2(n-1)p^{-1}\sum_{i=1}^{p}\|\Sigma_i\|\|S_i-(n-1)\Sigma_i\|\right)\\
&=\frac{4(n-1)^2}{p}p^{-1}\sum_{i=1}^{p}\operatorname{Var}\big(\|\Sigma_i\|\|S_i-(n-1)\Sigma_i\|\big)\\
&\le\frac{4(n-1)^2}{p}p^{-1}\sum_{i=1}^{p}E\big(\|\Sigma_i\|^2\|S_i-(n-1)\Sigma_i\|^2\big)\\
&=\frac{4(n-1)^2}{p}p^{-1}\sum_{i=1}^{p}\operatorname{tr}(\Sigma_i^2)\Big[E\operatorname{tr}(S_i^2)-2(n-1)E\operatorname{tr}(S_i\Sigma_i)+(n-1)^2\operatorname{tr}(\Sigma_i^2)\Big]\\
&=\frac{4(n-1)^2}{p}p^{-1}\sum_{i=1}^{p}\operatorname{tr}(\Sigma_i^2)\Big[n(n-1)\operatorname{tr}(\Sigma_i^2)+(n-1)(\operatorname{tr}\Sigma_i)^2-2(n-1)^2\operatorname{tr}(\Sigma_i^2)+(n-1)^2\operatorname{tr}(\Sigma_i^2)\Big]\\
&=\frac{4(n-1)^3}{p}p^{-1}\sum_{i=1}^{p}\big(\operatorname{tr}(\Sigma_i^2)\big)^2+\operatorname{tr}(\Sigma_i^2)(\operatorname{tr}\Sigma_i)^2\overset{p\to\infty}{\longrightarrow}0.
\end{aligned}
$$

By Markov's inequality, we have

$$
\left|p^{-1}\sum_{i=1}^{p}2(n-1)\Big[\operatorname{tr}(S_i\Sigma_i)-(n-1)\operatorname{tr}(\Sigma_i^2)\Big]\right|\overset{\text{prob}}{\longrightarrow}0\quad\text{as } p\to\infty. \tag{19}
$$

For (15), we have

$$
\begin{aligned}
&\sup_{\nu>q+1,\|\Psi\|\le\max_{1\le i\le p}\|S_i\|}\left|p^{-1}\sum_{i=1}^{p}\frac{2\big\langle\Psi-(\nu-q-1)\Sigma_i,S_i-(n-1)\Sigma_i\big\rangle}{(\nu+n-q-2)^2}\right|\\
&\le\sup_{\nu>q+1,\|\Psi\|\le\max_{1\le i\le p}\|S_i\|}\left|p^{-1}\sum_{i=1}^{p}\frac{2\big\langle\Psi,S_i-(n-1)\Sigma_i\big\rangle}{(\nu+n-q-2)^2}\right|\\
&\quad+\sup_{\nu>q+1}\left|p^{-1}\sum_{i=1}^{p}\frac{2(\nu-q-1)\big\langle\Sigma_i,S_i-(n-1)\Sigma_i\big\rangle}{(\nu+n-q-2)^2}\right|\\
&\le\sup_{\|\Psi\|\le\max_{1\le i\le p}\|S_i\|}\left|p^{-1}\sum_{i=1}^{p}2\big\langle\Psi,S_i-(n-1)\Sigma_i\big\rangle\right|+\left|p^{-1}\sum_{i=1}^{p}2\big\langle\Sigma_i,S_i-(n-1)\Sigma_i\big\rangle\right|.
\end{aligned}
$$

Since by assumption (i)

$$\begin{aligned}
&\operatorname{Var}\left(p^{-1}\sum_{i=1}^{p} 2\big\langle \Sigma_i, S_i-(n-1)\Sigma_i\big\rangle\right)\\
&= \frac{1}{p^2}\sum_{i=1}^{p} E\big[\langle \Sigma_i, S_i-(n-1)\Sigma_i\rangle^2\big]\\
&\le \frac{1}{p^2}\sum_{i=1}^{p} E\big[\|\Sigma_i\|^2\|S_i-(n-1)\Sigma_i\|^2\big]\\
&= \frac{1}{p^2}\sum_{i=1}^{p} \|\Sigma_i\|^2 E\big[\operatorname{tr}(S_i^2)-2(n-1)\operatorname{tr}(S_i\Sigma_i)+(n-1)^2\operatorname{tr}(\Sigma_i^2)\big]\\
&= \frac{1}{p^2}\sum_{i=1}^{p} \operatorname{tr}(\Sigma_i^2)\Big[(n-1)\operatorname{tr}(\Sigma_i^2)+(n-1)(\operatorname{tr}\Sigma_i)^2\Big]\\
&\le \frac{n-1}{p^2}\sum_{i=1}^{p}(\operatorname{tr}\Sigma_i)^4 \overset{p\to\infty}{\longrightarrow} 0,
\end{aligned}$$

by Markov's inequality, we have

$$\left|p^{-1}\sum_{i=1}^{p} 2\big\langle \Sigma_i, S_i-(n-1)\Sigma_i\big\rangle\right| \overset{\text{prob}}{\longrightarrow} 0 \quad \text{as } p\to\infty.$$

Similarly, we have

$$\sup_{\|\Psi\|\le \max_{1\le i\le p}\|S_i\|}\left|p^{-1}\sum_{i=1}^{p} 2\big\langle \Psi, S_i-(n-1)\Sigma_i\big\rangle\right| \le \frac{2}{p}\max_{1\le i\le p}\|S_i\|\left\|\sum_{i=1}^{p} S_i-(n-1)\Sigma_i\right\|$$

and

$$E\left[\frac{2}{p}\max_{1\le i\le p}\|S_i\|\left\|\sum_{i=1}^{p} S_i-(n-1)\Sigma_i\right\|\right]=o(1).$$

Hence, we have

$$\sup_{\nu>q+1,\|\Psi\|\le \max_{1\le i\le p}\|S_i\|}\left|p^{-1}\sum_{i=1}^{p}\frac{2\big\langle \Psi-(\nu-q-1)\Sigma_i, S_i-(n-1)\Sigma_i\big\rangle}{(\nu+n-q-2)^2}\right| \overset{\text{prob}}{\longrightarrow} 0 \quad \text{as } p\to\infty. \tag{20}$$

Combining (16), (17), (18), (19), and (20), we have

$$\sup_{\nu>q+1,\|\Psi\|\le \max_{1\le i\le p}\|S_i\|}\left|\mathrm{SURE}_2(\nu,\Psi)-L_2\big(\widehat{\boldsymbol{\Sigma}}^{\nu,\Psi},\boldsymbol{\Sigma}\big)\right| \overset{\text{prob}}{\longrightarrow} 0 \quad \text{as } p\to\infty. \tag{21}$$

The proof is concluded by (10) and (21). □

Theorem 4 *If assumptions (i), (ii), and (iii) in Theorem 3 hold, then*

$$\limsup_{p\to\infty}\left[R\Big((\widehat{\boldsymbol{M}}^{SURE},\widehat{\boldsymbol{\Sigma}}^{SURE}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-R\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)\right]\le 0.$$

Proof. Since

$$\begin{aligned}
&L\Big((\widehat{\boldsymbol{M}}^{\mathrm{SURE}},\widehat{\boldsymbol{\Sigma}}^{\mathrm{SURE}}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-L\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)\\
&=L\Big((\widehat{\boldsymbol{M}}^{\mathrm{SURE}},\widehat{\boldsymbol{\Sigma}}^{\mathrm{SURE}}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-\mathrm{SURE}(\widehat{\lambda}^{\mathrm{SURE}},\widehat{\mu}^{\mathrm{SURE}},\widehat{\Psi}^{\mathrm{SURE}},\widehat{\nu}^{\mathrm{SURE}})\\
&\quad+\mathrm{SURE}(\widehat{\lambda}^{\mathrm{SURE}},\widehat{\mu}^{\mathrm{SURE}},\widehat{\Psi}^{\mathrm{SURE}},\widehat{\nu}^{\mathrm{SURE}})-\mathrm{SURE}(\lambda,\mu,\Psi,\nu)\\
&\quad+\mathrm{SURE}(\lambda,\mu,\Psi,\nu)-L\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)\\
&\le\sup_{\lambda,\mu,\Psi,\nu}\left|L\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-\mathrm{SURE}(\lambda,\mu,\Psi,\nu)\right|\\
&\quad+0\\
&\quad+\sup_{\lambda,\mu,\Psi,\nu}\left|L\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-\mathrm{SURE}(\lambda,\mu,\Psi,\nu)\right|\\
&=2\sup_{\lambda,\mu,\Psi,\nu}\left|L\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-\mathrm{SURE}(\lambda,\mu,\Psi,\nu)\right|,
\end{aligned}$$

from Theorem 3, we have

$$\limsup_{p\to\infty}\left[L\Big((\widehat{\boldsymbol{M}}^{\mathrm{SURE}},\widehat{\boldsymbol{\Sigma}}^{\mathrm{SURE}}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-L\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)\right]\le 0.$$

Hence, by dominated convergence, we have

$$\begin{aligned}
&\limsup_{p\to\infty}\left[R\Big((\widehat{\boldsymbol{M}}^{\mathrm{SURE}},\widehat{\boldsymbol{\Sigma}}^{\mathrm{SURE}}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-R\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)\right]\\
&=\limsup_{p\to\infty}E\left[L\Big((\widehat{\boldsymbol{M}}^{\mathrm{SURE}},\widehat{\boldsymbol{\Sigma}}^{\mathrm{SURE}}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-L\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)\right]\\
&=E\left\{\limsup_{p\to\infty}\left[L\Big((\widehat{\boldsymbol{M}}^{\mathrm{SURE}},\widehat{\boldsymbol{\Sigma}}^{\mathrm{SURE}}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)-L\Big((\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu}),(\boldsymbol{M},\boldsymbol{\Sigma})\Big)\right]\right\}\\
&\le 0.
\end{aligned}$$

□

4 Implementation Details

Note that to find the shrinkage estimators $(\widehat{\boldsymbol{M}}^{\lambda,\mu},\widehat{\boldsymbol{\Sigma}}^{\Psi,\nu})$, we need to solve the optimization problem

$$\min_{\lambda,\mu,\Psi,\nu}\mathrm{SURE}(\lambda,\mu,\Psi,\nu)$$

which is a non-convex optimization problem. Hence the solution depends heavily on the initialization of the minimization algorithm. In this section, we provide a way to choose the initialization so that, in our experiments, the algorithm converges successfully in less than 10 iterations. We can compute the marginal expectations

$$E^{\mathrm{LE}}\big(\bar{X}_i^{\mathrm{LE}}\big) = E_{M_i}^{\mathrm{LE}}\Big[E_X^{\mathrm{LE}}\big(\bar{X}_i^{\mathrm{LE}}|M_i\big)\Big] = E_{M_i}^{\mathrm{LE}}[M_i] = \mu \tag{22}$$

$$\begin{aligned} Ed^2_{\mathrm{LE}}(\bar{X}_i^{\mathrm{LE}}, \mu) &= E_{\Sigma_i}\big\{E_{M_i}\big[E_X\big(d^2_{\mathrm{LE}}(\bar{X}_i^{\mathrm{LE}}, \mu)|M_i, \Sigma_i\big)|\Sigma_i\big]\big\} \\ &= E_{\Sigma_i}\Big[\Big(1/n + 1/\lambda\Big)\operatorname{tr}\Sigma_i\Big] = \left(\frac{1}{n} + \frac{1}{\lambda}\right)\frac{\operatorname{tr}\Psi}{\nu - q - 1} \end{aligned} \tag{23}$$

$$E(S_i) = E_{\Sigma_i}[E_{S_i}(S_i|\Sigma_i)] = E_{\Sigma_i}[(n-1)\Sigma_i] = \frac{n-1}{\nu - q - 1}\Psi \tag{24}$$

$$E(S_i^{-1}) = E_{\Sigma_i}[E_{S_i}(S_i^{-1}|\Sigma_i)] = E_{\Sigma_i}\left[\frac{\Sigma_i^{-1}}{n-q-2}\right] = \frac{\nu}{n-q-2}\Psi^{-1} \tag{25}$$

where $E^{\mathrm{LE}}(\cdot)$ denotes the Fréchet expectation with respect to the Log-Euclidean metric, i.e. $E^{\mathrm{LE}}(X) = \exp(E(\log X))$. Thus the hyperparameters can be written as

$$\begin{aligned} \mu &= E^{\mathrm{LE}}\big(\bar{X}_i^{\mathrm{LE}}\big) \\ \lambda &= \frac{nEd^2_{\mathrm{LE}}(\bar{X}_i^{\mathrm{LE}}, \mu)}{\frac{n}{n-1}E(\operatorname{tr} S_i) - Ed^2_{\mathrm{LE}}(\bar{X}_i^{\mathrm{LE}}, \mu)} \quad \text{(by (23) and (24))} \\ \nu &= \frac{q+1}{\frac{n-q-2}{q(n-1)}\operatorname{tr}(E(S_i)E(S_i^{-1})) - 1} + q + 1 \quad \text{(by (24) and (25))} \\ \Psi &= \frac{\nu - q - 1}{n-1}E(S_i) \quad \text{(by (24))} \end{aligned}$$

and an initialization of the hyperparameters can be obtained by replacing the (Fréchet) expectations with the corresponding sample (Fréchet) means, i.e.

$$\begin{aligned} \mu_0 &= \exp\left(p^{-1}\sum_{i=1}^{p}\log\bar{X}_i^{\mathrm{LE}}\right) \\ \lambda_0 &= \frac{np^{-1}\sum_{i=1}^{p} d^2_{\mathrm{LE}}(\bar{X}_i^{\mathrm{LE}}, \mu_0)}{\frac{n}{p(n-1)}\sum_{i=1}^{p}\operatorname{tr} S_i - p^{-1}\sum_{i=1}^{p} d^2_{\mathrm{LE}}(\bar{X}_i^{\mathrm{LE}}, \mu_0)} \\ \nu_0 &= \frac{q+1}{\frac{n-q-2}{p^2 q(n-1)}\operatorname{tr}\Big[(\sum_{i=1}^{p} S_i)(\sum_{i=1}^{p} S_i^{-1})\Big] - 1} + q + 1 \\ \Psi_0 &= \frac{\nu_0 - q - 1}{p(n-1)}\sum_{i=1}^{p} S_i. \end{aligned}$$

Note that these initial values can also be viewed as empirical Bayes estimates for μ, λ, ν, and Ψ obtained by matching moments. However these estimates do not possess the asymptotic optimality as stated in Theorem 3 and Theorem 4 since they are not obtained by minimizing an estimate of the risk function.

5 Data Descriptions

In this section, we provide the details for the data used in Section 4.2.

5.1 Resting State Functional MRI Connectivity Networks

In this experiment, we have three datasets, named ADHD200_CC200, PRURIM, and UCSF_MAC_PSP. Each contains brain connectivity matrices from different subjects, and these subjects are further categorized into different groups according to their medical conditions. These groups are summarized in Table 1. For example, in the 'Typically Developing' group of the 'ADHD200_CC200' dataset, there are 330 connectivity matrices (or correlation matrices), each of size 190×190. Our goal here is to estimate the group means (for $p = 7$ groups in total) using these three independent datasets. The baseline approach is to use the sample Fréchet means (FMs) from each group as the estimators. Our approach is to shrink these sample FMs to achieve better estimation performance. However, our theory only shrinks SPD matrices that are of the same size, and this is not the case for these datasets (we have three sizes: 190×190, 116×116, 27×27). Therefore we pick only a highly correlated sub-network of size $N \times N$ from each group instead.

Studies	Groups (Medical Condition)	# of patients	Matrix Size
ADHD200_CC200	Typically Developing	330	190×190
	ADHD-Combined	109	
	ADHD-Inattentive	74	
PRURIM	Healthy	15	116×116
	Psoriasis	14	
UCSF_MAC_PSP	Control	40	27×27
	Progressive Supranuclear Palsy	24	

Table 1: Specifics of the three dataset used in the rs-fMRI experiment.

The procedure for choosing sub-networks for each group is as follows. For each group, we first compute the mean connectivity network, and then apply a hierarchical clustering algorithm to choose the N nodes with the highest absolute correlation. For example, for the 'Typically Developing' group of the 'ADHD200_CC200' dataset, there are 190 nodes in the connectivity matrix. For any two nodes x, y, we measure the dissimilarity by $d(x, y) = 1 - |\text{cor}(x, y)|$, and with these dissimilarity measures, we build a dendrogram and select the N nodes that are closest to each other. Note that each node of the connectivity matrix represents a physical region in the brain, and the chosen sub-network is the correlation matrix of the highly correlated regions in the brain. Also, since this procedure for choosing sub-networks is applied in each group separately, the sub-networks from different groups do not have the same nodes/regions (even if they are from the same dataset). This does not pose a problem for our experiment as our theory assumes independence between groups.

6 Additional Simulation Studies

In this section, we present some additional simulation studies to answer two questions: (1) Is our estimator still better than the MLE if the independence assumption is violated? (2) How does our estimator perform compared to the estimator obtained by simply applying James-Stein shrinkage to the log-transformed SPD matrices? The two simple simulation studies in Sections 6.1 and 6.2 give positive answers to both questions.

6.1 Comparison of the Sure Estimator and the Ordinary James-Stein Estimator

As pointed out in the Introduction, one simple method for developing a shrinkage estimator for SPD matrices is to simply apply James-Stein shrinkage to the log-transformed SPD matrices (and use the matrix exponential map to obtain the shrunk SPD matrices). However this does not lead to an optimal shrinkage estimator. In this section, we present a simulation study comparing our proposed estimator and the simple method described above.

The settings for this simulation study are $N = 3$, $p = 100$, $n = 1$, $\Sigma_i = 0.3I_3$, and

$$M_i = \begin{bmatrix} 1 & \rho_i & \rho_i^2 \\ \rho_i & 1 & \rho_i \\ \rho_i^2 & \rho_i & 1 \end{bmatrix},$$

where $\rho_i = \rho + U_i$, $U_i \overset{\text{iid}}{\sim} \text{Unif}(-0.01, 0.01)$, and $-0.8 \leq \rho \leq 0.8$. The notation follows that in Section 3 of the main paper. We compare three estimators: the MLE, the SURE estimator, and the estimator based on James-Stein shrinkage in the log-domain. The results are shown in Figure 1. Average loss is an average over 100 repetitions.

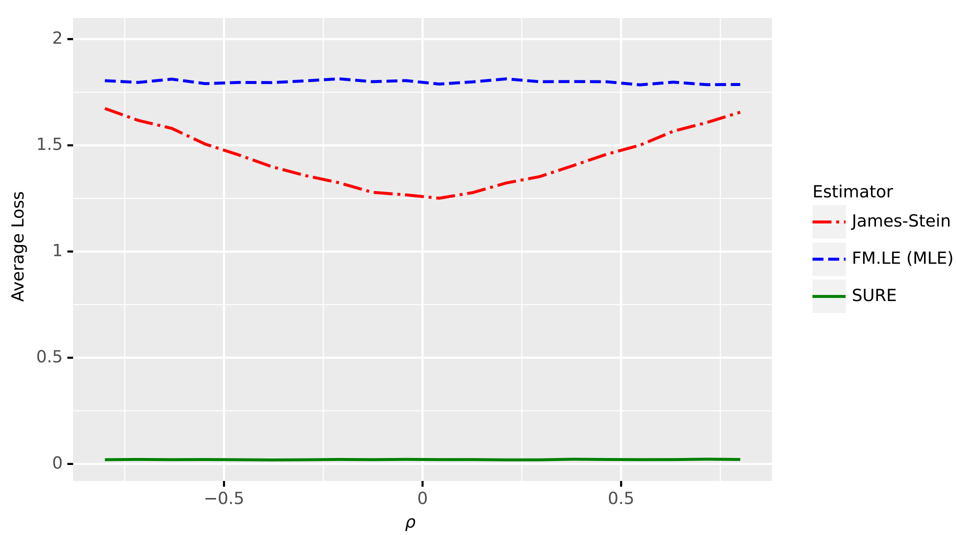

Figure 1: Average loss for each of the three estimators as ρ varies.

Note that when $\rho = 0$, $M_i \approx I_3$ and $\log I_3$ is the zero matrix. When $\rho = 0$ the James-Stein based shrinkage estimator yields maximal improvement over the MLE. As $|\rho|$ increases, the improvement becomes is less significant. This is consistent with the classical behavior of the James-Stein estimator. In contrast, the SURE estimator achieves optimal improvement regardless of the value of ρ. Figure 1 conveys two key messages. (1) Because the James-Stein estimator shrinks the MLE towards a fixed target (the identity matrix in this case), it achieves a significant improvement only when the true parameter is close to the shrinkage target. In contrast, the proposed SURE estimator shrinks the MLE towards a target determined from the data, and hence achieves significant improvement regardless of what is the true parameter. (2) In this particular simulation, the loss for the SURE estimator is nearly zero. The reason is that the variation in ρ_i is quite small ($\rho - 0.01 < \rho_i < \rho + 0.01$) and the shrinkage target determined from the data will be reliable. If, for example, $\rho_i \overset{\text{iid}}{\sim} \text{Unif}(-1, 1)$, then the shrinkage target determined from the data will be much less reliable, and therefore we will observe a larger loss for the SURE estimator.

6.2 Dependent Data

In this section, we discuss whether the proposed estimator is robust to the independence assumption, i.e. whether our estimator still dominates the MLE if the independence assumption is violated. In terms of our main application, DTI analysis, the independence assumption is actually unrealistic, made in order to simplify the mathematical analysis. In DTI, estimates for neighboring voxels will never be independent, and therefore we would like to know how the magnitude of the dependence affects the performance of our estimator.

We created an artificial dependence structure by computing the moving k-average (FM) of an array of independent SPD matrices, i.e., if $X = [X_1, \ldots, X_{p+k-1}]$ is an array of independent SPD matrices, then the moving k-average of X is $X^{(k)} = [\bar{X}^{(1)}, \ldots, \bar{X}^{(p)}]$ where $\bar{X}^{(i)} = \text{FM}(X_i, \ldots, X_{i+k-1})$. If $k = 1$ then $X^{(1)} = X$ and the entries in this array are independent, but when $k > 1$ there is dependence and this dependence increases when k increases. In this simulation study, we fixed the other parameters: $N = 3$, $p = 100$, $\lambda = 50$, $\mu = I_3$, $\Psi = I_6$, $\nu = 30$, and $n = 10$. Again, we compared with the same estimators as in Section 4.1. The results are shown in Figure 2. From the figure, we can clearly see that as k increases, the improvement is less significant, and in the extreme case where X contains p copies of an SPD matrix, there is no benefit in shrinking the MLE. But we also see that even when there is mild dependence within the observation, our shrinkage estimator is still better than the MLE.

7 Remarks on the Resampling Scheme in Section 4.2.1

Here we discuss the resampling scheme that is used in Section 4.2.1 and again in Section 4.2.2. In particular, we explain why we resample with replacement, as opposed to without replacement, and we mention another natural resampling scheme and explain why we do not use it.

Our explanations are in the context of the Parkinson's group in the illustration in Section 4.2.1. For patient j in that group, we have estimates $X_{i,j}$, $i = 1, \ldots, 78$, corresponding to the $44 + 34$ voxels. These are the estimates given by the software used in DTI. Also, let $X_{i,\cdot}$ be the Fréchet mean of X_{ij}, $j = 1, \ldots, 50$. Let F denote the distribution of

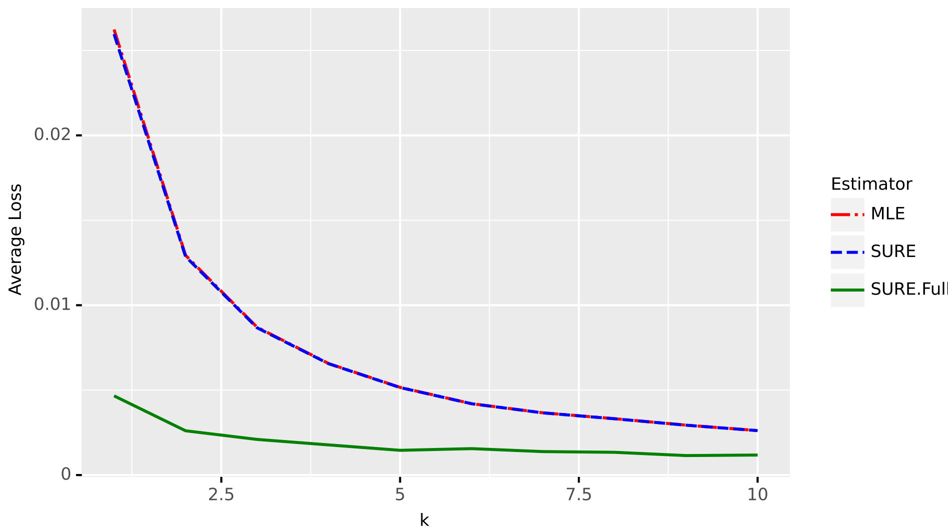

Figure 2: Average loss for each of the three estimators with varying k.

$(X_{1,1}, X_{2,1}, \ldots, X_{78,1})$. The 50 78-tuples $(X_{1,1}, X_{2,1}, \ldots, X_{78,1}), \ldots, (X_{1,50}, X_{2,50}, \ldots, X_{78,50})$ are iid draws from F. For each voxel i ($i = 1, \ldots, 78$) there is a true SPD matrix, M_i, particular to voxel i. This M_i may be defined as $X_{i,\cdot}$ in a hypothetical population that is not the 50 Parkinson's patients in this study, but rather is an infinite population from which the 50 Parkinson's patients are a random sample.

Now, consider any one of the three estimation methods FM.LE, SURE-FM, or SURE.Full-FM, and for concreteness consider FM.LE. This method produces an estimate $(\widehat{M}_1, \ldots, \widehat{M}_{78})$, based on the 50 Parkinson's patients. To evaluate the risk of this method, we need to consider the loss $L\big((\widehat{M}_1, \ldots, \widehat{M}_{78}), (M_1, \ldots, M_{78})\big)$, and take the expected value of this loss, i.e. find $E\big[L\big((\widehat{M}_1, \ldots, \widehat{M}_{78}), (M_1, \ldots, M_{78})\big)\big]$, where in the expected value the raw data used to calculate $(\widehat{M}_1, \ldots, \widehat{M}_{78})$ are a random sample (meaning iid sample) from F. This cannot be done analytically, so we do it via a simulation. Now F is unknown, so we estimate it by the empirical distribution, which gives mass 1/50 to each of the 50 points $(X_{1,1}, X_{2,1}, \ldots, X_{78,1}), \ldots, (X_{1,50}, X_{2,50}, \ldots, X_{78,50})$. This is exactly what we did in Section 4.2.1, and we note that iid sampling from the empirical distribution means sampling with replacement from the 50 78-tuples.

Now F is a distribution on 78-tuples of SPD matrices, and our model stipulates that the distribution of component i is a log-normal. So it is natural to ask why are we sampling from the empirical distribution instead of sampling from the log-normals with estimated parameters. The former corresponds to the ordinary bootstrap and the latter to the parametric bootstrap (see Section 6.5 of Efron & Tibshirani (1993)). The reason we do not use the parametric bootstrap is that for a given patient the estimates for M_i and M_j may be correlated, i.e. there is a patient effect, where for some patients the estimates of $M_1, \ldots, M_{78}$ all have large eigenvalues, while for others the estimates all have small eigenvalues. By sampling from the empirical distribution, we preserve this structure.

References

Efron, B. & Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman and Hall New York.

Gupta, A. K. & Nagar, D. K. (2000), *Matrix Variate Distributions*, Chapman and Hall/CRC.

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons.

Xie, X., Kou, S. C. & Brown, L. D. (2012), 'SURE estimates for a heteroscedastic hierarchical model', *Journal of the American Statistical Association* **107**(500), 1465–1479.