Double shrinkage priors for a normal mean matrix

Takeru Matsuda[1,2], Fumiyasu Komaki[1,2], and William E. Strawderman[3]

[1]The University of Tokyo
[2]RIKEN Center for Brain Science
[3]Rutgers University

Abstract

We consider estimation of a normal mean matrix under the Frobenius loss. Motivated by the Efron–Morris estimator, a generalization of Stein's prior has been recently developed, which is superharmonic and shrinks the singular values towards zero. The generalized Bayes estimator with respect to this prior is minimax and dominates the maximum likelihood estimator. However, here we show that it is inadmissible by using Brown's condition. Then, we develop two types of priors that provide improved generalized Bayes estimators and examine their performance numerically. The proposed priors attain risk reduction by adding scalar shrinkage or column-wise shrinkage to singular value shrinkage. Parallel results for Bayesian predictive densities are also given.

1 Introduction

Suppose that we have independent matrix observations $Y^{(1)}, \dots, Y^{(N)} \in \mathbb{R}^{n \times p}$ whose entries are independent normal random variables $Y_{ij}^{(t)} \sim \mathrm{N}(M_{ij}, 1)$, where $M \in \mathbb{R}^{n \times p}$ is an unknown mean matrix. In the notation of Gupta and Nagar (2000), this is expressed as $Y^{(t)} \sim \mathrm{N}_{n,p}(M, I_n, I_p)$ for $t = 1, \dots, N$, where I_k denotes the k-dimensional identity matrix. We consider estimation of M under the Frobenius loss

$$l(M, \hat{M}) = \|\hat{M} - M\|_{\mathrm{F}}^2 = \sum_{a=1}^{n} \sum_{i=1}^{p} (\hat{M}_{ai} - M_{ai})^2.$$

By sufficiency reduction, it suffices to consider the average $Y = (Y^{(1)} + \cdots + Y^{(N)})/N \sim \mathrm{N}(M, I_n, N^{-1} I_p)$ in estimation of M. We assume $n - p - 1 > 0$ in the following. Note that vectorization reduces this problem to estimation of a normal mean vector $\mathrm{vec}(M)$ from $\mathrm{vec}(Y) \sim \mathrm{N}_{np}(\mathrm{vec}(M), N^{-1} I_{np})$ under the quadratic loss, which has been well studied in shrinkage estimation theory (Fourdrinier et al., 2018).

Efron and Morris (1972) proposed an empirical Bayes estimator:

$$\hat{M}_{\mathrm{EM}} = Y \left(I_p - \frac{n-p-1}{N} (Y^\top Y)^{-1} \right). \tag{1}$$

This estimator can be viewed as a generalization of the James–Stein estimator ($p = 1$) for a normal mean vector. Efron and Morris (1972) showed that $\hat{M}_{\mathrm{EM}}$ is minimax and dominates the maximum likelihood estimator $\hat{M} = Y$ under the Frobenius loss. Let $Y = U\Lambda V^\top$, $U \in \mathbb{R}^{n\times p}$, $V \in \mathbb{R}^{p\times p}$, $\Lambda = \mathrm{diag}(\sigma_1(Y), \dots, \sigma_p(Y))$ be the singular value decomposition of Y, where $U^\top U = V^\top V = I_p$ and $\sigma_1(Y) \geq \cdots \geq \sigma_p(Y) \geq 0$ are the singular values of Y. Stein (1974) pointed out that $\hat{M}_{\mathrm{EM}}$ does not change the singular vectors but shrinks the singular values of Y towards zero:

$$\hat{M}_{\mathrm{EM}} = U\hat{\Lambda}_{\mathrm{EM}}V^\top, \quad \hat{\Lambda}_{\mathrm{EM}} = \mathrm{diag}(\sigma_1(\hat{M}_{\mathrm{EM}}), \dots, \sigma_p(\hat{M}_{\mathrm{EM}})),$$

where

$$\sigma_i(\hat{M}_{\mathrm{EM}}) = \left(1 - \frac{n-p-1}{N\sigma_i(Y)^2}\right)\sigma_i(Y), \quad i = 1, \dots, p.$$

See Tsukuma and Kubokawa (2020); Yuasa and Kubokawa (2023a,b) for recent developments around the Efron–Morris estimator.

As a Bayesian counterpart of $\hat{M}_{\mathrm{EM}}$, Matsuda and Komaki (2015) proposed a singular value shrinkage prior

$$\pi_{\mathrm{SVS}}(M) = \det(M^\top M)^{-(n-p-1)/2}, \tag{2}$$

and showed that the generalized Bayes estimator $\hat{M}_{\mathrm{SVS}}$ with respect to π_{SVS} dominates the maximum likelihood estimator $\hat{M} = Y$ under the Frobenius loss. This prior can be viewed as a generalization of Stein's prior $\pi(\mu) = \|\mu\|^{2-n}$ for a normal mean vector μ ($p = 1$) by Stein (1974). Similarly to $\hat{M}_{\mathrm{EM}}$ in (1) , $\hat{M}_{\mathrm{SVS}}$ shrinks the singular values towards zero. Thus, it works well when the true matrix is close to low-rank. See Matsuda and Strawderman (2022) and Matsuda (2023) for details on the risk behavior of $\hat{M}_{\mathrm{EM}}$ and $\hat{M}_{\mathrm{SVS}}$.

In this paper, we show that the generalized Bayes estimator with respect to the singular value shrinkage prior π_{SVS} in (2) is inadmissible under the Frobenius loss. Then, we develop two types of priors that provide improved generalized Bayes estimators asymptotically. The first type adds scalar shrinkage while the second type adds column-wise shrinkage. We conduct numerical experiments and confirm the effectiveness of the proposed priors in finite samples. We also provide parallel results for Bayesian prediction as well as a similar improvement of the blockwise Stein prior, which was conjectured by Brown and Zhao (2009).

This paper is organized as follows. In Section 2, we prove the inadmissibility of the generalized Bayes estimator with respect to the singular value shrinkage prior π_{SVS} in (2). In Sections 3 and 4, we provide two types of priors that asymptotically dominate the singular value shrinkage prior π_{SVS} in (2) by adding scalar or column-wise shrinkage, respectively. Numerical results are also given. In Section 5, we provide parallel results for Bayesian prediction. Technical details and similar results for the blockwise Stein prior are given in the Appendix.

2 Inadmissibility of the singular value shrinkage prior

Here, we show that the generalized Bayes estimator with respect to the singular value shrinkage prior π_{SVS} in (2) is inadmissible under the Frobenius loss. Since N does not affect admissibility results, we fix $N = 1$ for convenience in this section.

For estimation of a normal mean vector under the quadratic loss, Brown (1971) derived the following sufficient condition for inadmissibility of generalized Bayes estimators.

Lemma 2.1. *(Brown, 1971) In estimation of θ from $Y \sim \mathrm{N}_d(\theta, I_d)$ under the quadratic loss, the generalized Bayes estimator of θ with respect to a prior $\pi(\theta)$ is inadmissible if*

$$\int_c^\infty r^{1-d} \underline{m}(r) \mathrm{d}r < \infty$$

for some $c > 0$, where

$$\underline{m}(r) = \int \frac{1}{m_\pi(y)} \mathrm{d}U_r(y),$$

$$m_\pi(y) = \int p(y \mid \theta)\pi(\theta)\mathrm{d}\theta,$$

$$p(y \mid \theta) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|y-\theta\|^2}{2}\right),$$

and U_r is the uniform measure on the sphere of radius r in $\mathbb{R}^d$.

After vectorization, estimation of a normal mean matrix M from $Y \sim \mathrm{N}_{n,p}(M, I_n, I_p)$ under the Frobenius loss reduces to estimation of a normal mean vector $\mathrm{vec}(M)$ from $\mathrm{vec}(Y) \sim \mathrm{N}_{np}(\mathrm{vec}(M), I_{np})$ under the quadratic loss. Then, by using Brown's condition in Lemma 2.1, we obtain the following.

Theorem 2.1. *When $p \geq 2$, the generalized Bayes estimator with respect to π_{SVS} in (2) is inadmissible under the Frobenius loss.*

Proof. From $n-p-1>0$ and the AM-GM inequality

$$\left(\prod_{i=1}^p \sigma_i(M)^2\right)^{1/p} \leq \frac{1}{p}\sum_{i=1}^p \sigma_i(M)^2,$$

we have

$$\begin{aligned} \pi_{\mathrm{SVS}}(M) = \left(\prod_{i=1}^p \sigma_i(M)^2\right)^{-(n-p-1)/2} &\geq \left(\frac{1}{p}\sum_{i=1}^p \sigma_i(M)^2\right)^{-p(n-p-1)/2} \\ &= A_{n,p}\|M\|_{\mathrm{F}}^{-p(n-p-1)}, \end{aligned}$$

where $A_{n,p} = p^{p(n-p-1)/2}$. Therefore,

$$\begin{aligned} m_{\mathrm{SVS}}(Y) &= \int \pi_{\mathrm{SVS}}(M) p(Y \mid M)\mathrm{d}M \\ &\geq A_{n,p}\int \|M\|_{\mathrm{F}}^{-p(n-p-1)} p(Y \mid M)\mathrm{d}M \\ &= A_{n,p}\mathrm{E}[\|Y+Z\|_{\mathrm{F}}^{-p(n-p-1)}] \\ &\geq A_{n,p}\mathrm{E}[(\|Y\|_{\mathrm{F}}+\|Z\|_{\mathrm{F}})^{-p(n-p-1)}], \end{aligned} \tag{3}$$

where $Z = M - Y \sim \mathrm{N}_{n,p}(O, I_n, I_p)$ and we used the triangle inequality. As $\|Y\|_{\mathrm{F}} \to \infty$,

$$\|Y\|_{\mathrm{F}}^{p(n-p-1)}\mathrm{E}[(\|Y\|_{\mathrm{F}}+\|Z\|_{\mathrm{F}})^{-p(n-p-1)}] = \mathrm{E}\left[\left(1+\frac{\|Z\|_{\mathrm{F}}}{\|Y\|_{\mathrm{F}}}\right)^{-p(n-p-1)}\right] \to 1,$$

which yields

$$\mathrm{E}[(\|Y\|_\mathrm{F}+\|Z\|_\mathrm{F})^{-p(n-p-1)}]=O(\|Y\|_\mathrm{F}^{-p(n-p-1)}). \tag{4}$$

Now, we apply Lemma 2.1 by noting that estimation of a normal mean matrix M from $Y \sim \mathrm{N}_{n,p}(M, I_n, I_p)$ under the Frobenius loss is equivalent to estimation of a normal mean vector $\mathrm{vec}(M)$ from $\mathrm{vec}(Y) \sim \mathrm{N}_{np}(\mathrm{vec}(M), I_{np})$ under the quadratic loss. Let U_r be the uniform measure on the sphere of radius r in $\mathbb{R}^{n\times p}$, where the Frobenius norm is adopted for radius. Then, from (3) and (4),

$$\underline{m}_{\mathrm{SVS}}(r)=\int \frac{1}{m_{\mathrm{SVS}}(Y)}\mathrm{d}U_r(Y) \le Cr^{p(n-p-1)}$$

for some constant C. Therefore, since $-p^2-p+1<-1$ when $p\ge 2$,

$$\int_1^\infty r^{1-np}\underline{m}_{\mathrm{SVS}}(r)\mathrm{d}r \le C\int_1^\infty r^{-p^2-p+1}\mathrm{d}r<\infty.$$

From Lemma 2.1, it implies the inadmissibility of the generelized Bayes estimator with respect to π_{SVS} under the Frobenius loss. $\square$

3 Improvement by additional scalar shrinkage

Here, motivated by the result of Efron and Morris (1976), we develop a class of priors for which the generalized Bayes estimators asymptotically dominate that with respect to the singular value shrinkage prior π_{SVS} in (2). Efron and Morris (1976) proved that the estimator

$$\hat{M}_{\mathrm{MEM}}=Y\left(I_p-\frac{n-p-1}{N}(Y^\top Y)^{-1}-\frac{p^2+p-2}{N\|Y\|_\mathrm{F}^2}I_p\right) \tag{5}$$

dominates $\hat{M}_{\mathrm{EM}}$ in (1) under the Frobenius loss. This estimator shrinks the singular values of Y more strongly than $\hat{M}_{\mathrm{EM}}$:

$$\hat{M}_{\mathrm{MEM}}=U\hat{\Lambda}_{\mathrm{MEM}}V^\top, \quad \hat{\Lambda}_{\mathrm{MEM}}=\mathrm{diag}(\sigma_1(\hat{M}_{\mathrm{MEM}}),\dots,\sigma_p(\hat{M}_{\mathrm{MEM}})),$$

where

$$\sigma_i(\hat{M}_{\mathrm{MEM}})=\left(1-\frac{n-p-1}{N\sigma_i(Y)^2}-\frac{p^2+p-2}{N\|Y\|_\mathrm{F}^2}\right)\sigma_i(Y), \quad i=1,\dots,p. \tag{6}$$

In other words, $\hat{M}_{\mathrm{MEM}}$ adds scalar shrinkage to $\hat{M}_{\mathrm{EM}}$. Konno (1990, 1991) showed corresponding results in the unknown covariance setting. By extending these results, Tsukuma and Kubokawa (2007) derived a general method for improving matrix mean estimators by adding scalar shrinkage.

Motivated by $\hat{M}_{\mathrm{MEM}}$ in (5), we construct priors by adding scalar shrinkage to π_{SVS} in (2):

$$\pi_{\mathrm{MSVS1}}(M)=\pi_{\mathrm{SVS}}(M)\|M\|_\mathrm{F}^{-\gamma}, \tag{7}$$

where $\gamma\ge 0$. Note that Tsukuma and Kubokawa (2017) studied this type of prior in the context of Bayesian prediction. Let

$$m_{\mathrm{MSVS1}}(Y)=\int p(Y\mid M)\pi_{\mathrm{MSVS1}}(M)\mathrm{d}M$$

be the marginal density of Y under the prior $\pi_{\mathrm{MSVS1}}(M)$.

Lemma 3.1. *If $0 \le \gamma < p^2 + p$, then $m_{\mathrm{MSVS1}}(Y) < \infty$ for every Y.*

Proof. Since $m_{\mathrm{MSVS1}}(Y)$ is interpreted as the expectation of $\pi_{\mathrm{MSVS1}}(M)$ under $M \sim \mathrm{N}_{n,p}(Y, I_n, I_p)$, it suffices to show that $\pi_{\mathrm{MSVS1}}(M)$ is locally integrable at every M.

First, consider $M \neq O$. Since

$$m_{\mathrm{SVS}}(Y) = \int \pi_{\mathrm{SVS}}(M) p(Y \mid M) \mathrm{d}M < \infty$$

for every Y from Lemma 1 of Matsuda and Komaki (2015), $\pi_{\mathrm{SVS}}(M)$ is locally integrable at M. Also, $\|M\|_{\mathrm{F}} > c$ for some $c > 0$ in a neighborhood of M. Thus, $\pi_{\mathrm{MSVS1}}(M) = \pi_{\mathrm{SVS}}(M) \|M\|_{\mathrm{F}}^{-\gamma}$ is locally integrable at M if $\gamma \ge 0$.

Next, consider $M = O$ and take its neighborhood $A = \{M \mid \|M\|_{\mathrm{F}} \le \varepsilon\}$ for $\varepsilon > 0$. To evaluate the integral on A, we use the variable transformation from M to (r, U), where $r = \|M\|_{\mathrm{F}}$ and $U = M/r$ so that $M = rU$. We have $\mathrm{d}M = r^{np-1}\mathrm{d}r\mathrm{d}U$. Also, from $\det(M^\top M) = r^{2p} \det(U^\top U)$,

$$\pi_{\mathrm{MSVS1}}(M) = r^{-p(n-p-1)-\gamma} \det(U^\top U)^{-(n-p-1)/2}.$$

Thus,

$$\begin{aligned}
&\int_A \pi_{\mathrm{MSVS1}}(M) \mathrm{d}M \\
&= \int_0^\varepsilon r^{-p(n-p-1)-\gamma+np-1} \mathrm{d}r \int \det(U^\top U)^{-(n-p-1)/2} \mathrm{d}U \\
&= \int_0^\varepsilon r^{p^2+p-\gamma-1} \mathrm{d}r \int \det(U^\top U)^{-(n-p-1)/2} \mathrm{d}U.
\end{aligned}$$

The integral with respect to r is finite if $p^2 + p - \gamma - 1 > -1$, which is equivalent to $\gamma < p^2 + p$. The integral with respect to U is finite due to the local integrability of π_{SVS}, which corresponds to $\gamma = 0$, at $M = O$. Therefore, $\pi_{\mathrm{MSVS1}}(M)$ is locally integrable at $M = O$ if $0 \le \gamma < p^2 + p$.

Hence, $\pi_{\mathrm{MSVS1}}(M)$ is locally integrable at every M if $0 \le \gamma < p^2 + p$. $\square$

From Lemma 3.1, the generalized Bayes estimator with respect to π_{MSVS1} is well-defined when $0 \le \gamma < p^2 + p$. We denote it by $\hat{M}_{\mathrm{MSVS1}}$.

Theorem 3.1. *For every M,*

$$N^2 (\mathrm{E}_M[\|\hat{M}_{\mathrm{MSVS1}} - M\|_{\mathrm{F}}^2] - \mathrm{E}_M[\|\hat{M}_{\mathrm{SVS}} - M\|_{\mathrm{F}}^2]) \to \frac{\gamma(\gamma - 2p^2 - 2p + 4)}{\mathrm{tr}(M^\top M)} \tag{8}$$

as $N \to \infty$. Therefore, if $p \ge 2$ and $0 < \gamma < p^2 + p$, then the generalized Bayes estimator with respect to π_{MSVS1} in (7) *asymptotically dominates that with respect to π_{SVS} in* (2) *under the Frobenius loss.*

Proof. Let $K = M^\top M$ and K^{ij} be the (i,j)th entry of K^{-1}. From

$$\frac{\partial K_{jk}}{\partial M_{ai}} = \delta_{ik} M_{aj} + \delta_{ij} M_{ak}, \quad \frac{\partial}{\partial K_{ij}} \det K = K^{ij} \det K, \tag{9}$$

we have

$$\frac{\partial}{\partial M_{ai}} \det K = \sum_{j,k} \frac{\partial K_{jk}}{\partial M_{ai}} \frac{\partial}{\partial K_{jk}} \det K = 2 \sum_j M_{aj} K^{ij} \det K.$$

Therefore,

$$\frac{\partial}{\partial M_{ai}} \log \pi_{\mathrm{SVS}}(M) = -(n-p-1) \sum_j M_{aj} K^{ij}. \tag{10}$$

Let $\pi_{\mathrm{S}}(M) = \|M\|_{\mathrm{F}}^{-\gamma} = (\mathrm{tr} K)^{-\gamma/2}$. Since

$$\frac{\partial}{\partial M_{ai}} \mathrm{tr} K = 2 M_{ai}$$

from (9), we have

$$\frac{\partial}{\partial M_{ai}} \log \pi_{\mathrm{S}}(M) = -\gamma M_{ai} (\mathrm{tr} K)^{-1}, \tag{11}$$

$$\frac{\partial^2}{\partial M_{ai}^2} \log \pi_{\mathrm{S}}(M) = -\gamma (\mathrm{tr} K - 2 M_{ai}^2)(\mathrm{tr} K)^{-2}. \tag{12}$$

By using (10), (11), and (12), we obtain

$$\begin{aligned}
\mathrm{tr}(\widetilde{\nabla} \log \pi_{\mathrm{SVS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{S}}(M)) &= \gamma p (n-p-1)(\mathrm{tr} K)^{-1}, \\
\mathrm{tr}(\widetilde{\nabla} \log \pi_{\mathrm{S}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{S}}(M)) &= \gamma^2 (\mathrm{tr} K)^{-1}, \\
\mathrm{tr}(\widetilde{\Delta} \log \pi_{\mathrm{S}}(M)) &= -\gamma (np-2)(\mathrm{tr} K)^{-1},
\end{aligned}$$

where we used the matrix derivative notations (28) and (29). Therefore, from Lemma A.2,

$$\begin{aligned}
&\mathrm{E}_M[\|\hat{M}_{\mathrm{MSVS1}} - M\|_{\mathrm{F}}^2] - \mathrm{E}_M[\|\hat{M}_{\mathrm{SVS}} - M\|_{\mathrm{F}}^2] \\
=& \frac{1}{N^2} \mathrm{tr}(2 \widetilde{\nabla} \log \pi_{\mathrm{SVS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{S}}(M) + \widetilde{\nabla} \log \pi_{\mathrm{S}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{S}}(M) + 2 \widetilde{\Delta} \log \pi_{\mathrm{S}}(M)) \\
&+ o(N^{-2}) \\
=& \frac{1}{N^2} \gamma (\gamma - 2p^2 - 2p + 4)(\mathrm{tr} K)^{-1} + o(N^{-2}).
\end{aligned}$$

Hence, we obtain (8). □

From (8), the choice $\gamma = p^2 + p - 2$ attains the minimum risk among $0 < \gamma < p^2 + p$. Note that $p^2 + p - 2$ also appears in the singular value decomposition form of the modified Efron–Morris estimator $\hat{M}_{\mathrm{MEM}}$ in (6).

Now, we examine the performance of π_{MSVS1} in (7) by Monte Carlo simulation. Figure 1 plots the Frobenius risk of generalized Bayes estimators with respect to π_{MSVS1} in (7) with $\gamma = p^2 + p - 2$, π_{SVS} in (2) and $\pi_{\mathrm{S}}(M) = \|M\|_{\mathrm{F}}^{2-np}$, which is Stein's prior on the vectorization of M, for $n = 10$, $p = 3$ and $N = 1$. We computed the generalized Bayes estimators by using the random-walk Metropolis–Hastings algorithm with Gaussian proposal of variance 0.1. Note that the Frobenius risk of these estimators depends only on the singular values of M due to the orthogonal invariance. Similarly to the Efron–Morris estimator and π_{SVS}, π_{MSVS1} works well when M is close to low-rank. Also, π_{MSVS1} attains large risk reduction when M is close to the zero matrix like π_{S}. Thus, π_{MSVS1} has the best of both worlds. Figure 2 plots the Frobenius risk for $n = 10$, $p = 3$ and $N = 10$, computed by the random walk Metropolis–Hastings algorithm with proposal variance 0.005. The risk behavior is similar to Figure 1. Figure 3 plots the Frobenius risk for $n = 20$, $p = 3$ and $N = 2$, computed by the random walk Metropolis–Hastings algorithm with proposal variance 0.01. Again, the risk behavior is similar to Figure 1. Note that the value of $np/N = 30$ is the same with Figure 1.

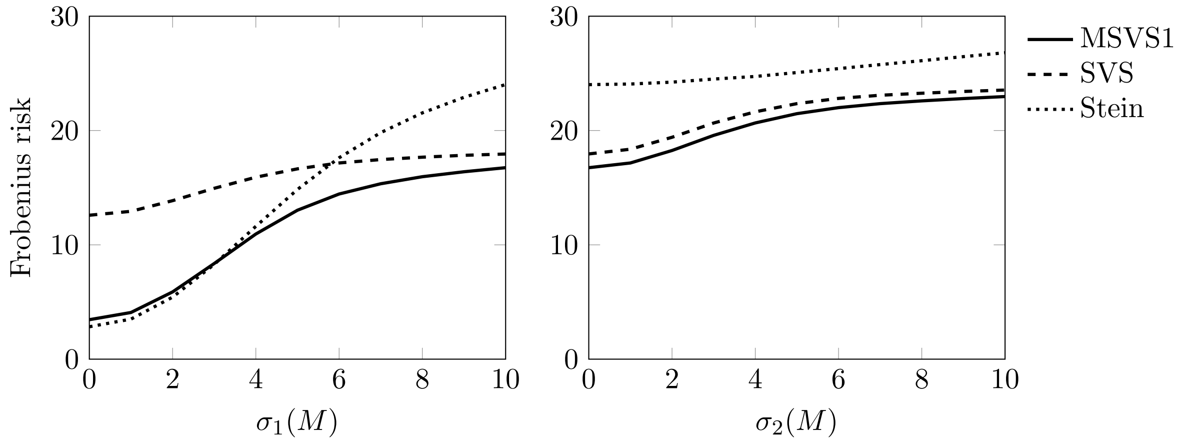

Figure 1: Frobenius risk of generalized Bayes estimators for $n = 10$, $p = 3$ and $N = 1$. Left: $\sigma_2 = \sigma_3 = 0$. Right: $\sigma_1 = 10$, $\sigma_3 = 0$. solid: π_{MSVS1} with $\gamma = p^2 + p - 2$, dashed: π_{SVS}, dotted: Stein's prior π_{S}, Note that the minimax risk is $np/N = 30$.

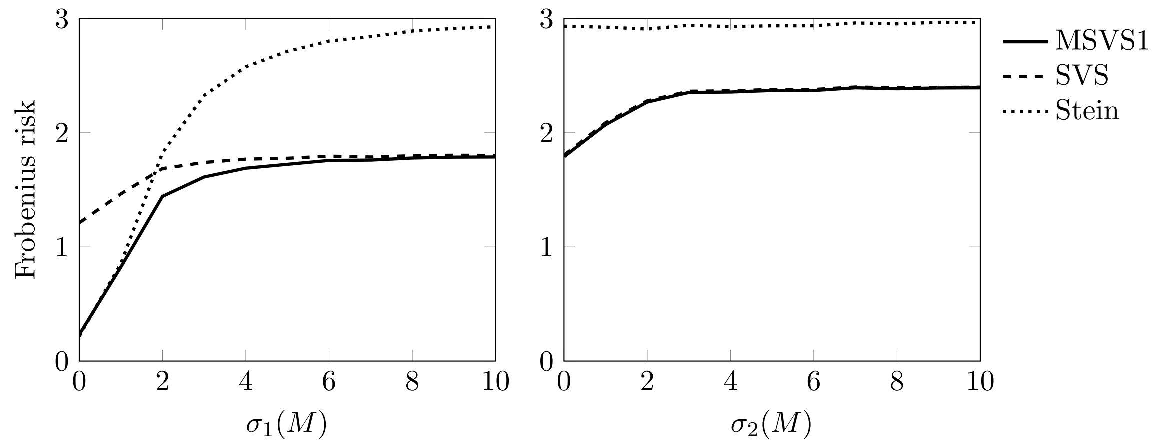

Figure 2: Frobenius risk of generalized Bayes estimators for $n = 10$, $p = 3$ and $N = 10$. Left: $\sigma_2 = \sigma_3 = 0$. Right: $\sigma_1 = 10$, $\sigma_3 = 0$. solid: π_{MSVS1} with $\gamma = p^2 + p - 2$, dashed: π_{SVS}, dotted: Stein's prior π_{S}, Note that the minimax risk is $np/N = 3$.

Improvement by additional scalar shrinkage holds even under the matrix quadratic loss (Matsuda and Strawderman, 2022; Matsuda, 2024).

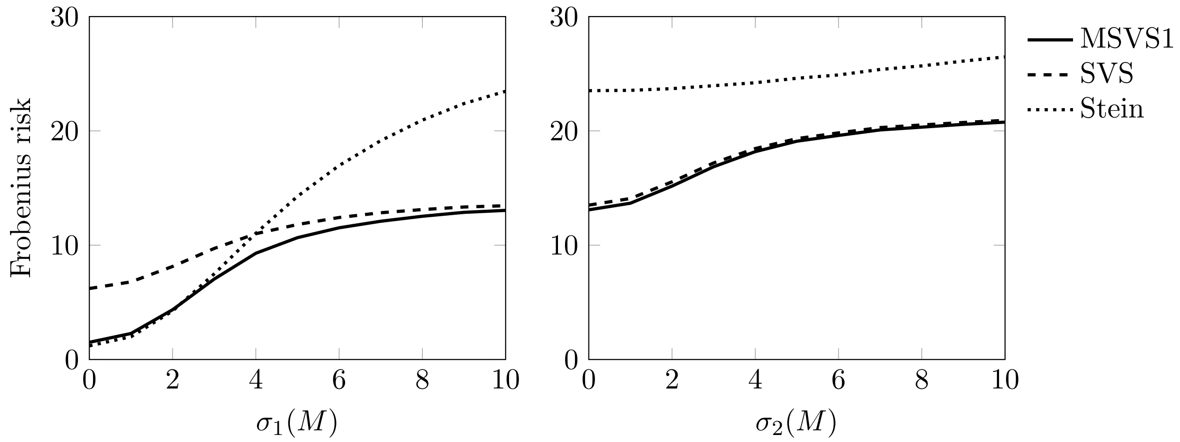

Figure 3: Frobenius risk of generalized Bayes estimators for $n = 20$, $p = 3$ and $N = 2$. Left: $\sigma_2 = \sigma_3 = 0$. Right: $\sigma_1 = 10$, $\sigma_3 = 0$. solid: π_{MSVS1} with $\gamma = p^2 + p - 2$, dashed: π_{SVS}, dotted: Stein's prior π_{S}, Note that the minimax risk is $np/N = 30$.

Theorem 3.2. *For every M,*

$$\begin{aligned} &N^2(\mathrm{E}_M[(\hat{M}_{\mathrm{MSVS1}} - M)^\top(\hat{M}_{\mathrm{MSVS1}} - M)] - \mathrm{E}_M[(\hat{M}_{\mathrm{SVS}} - M)^\top(\hat{M}_{\mathrm{SVS}} - M)]) \\ &\qquad \to \gamma(\mathrm{tr}K)^{-2}(-2(p+1)(\mathrm{tr}K)I_p + (\gamma+4)K) \end{aligned} \tag{13}$$

as $N \to \infty$. Therefore, if $p \geq 2$ and $0 < \gamma < 2p - 2$, then the generalized Bayes estimator with respect to π_{MSVS1} in (7) *asymptotically dominates that with respect to π_{SVS} in* (2) *under the matrix quadratic loss.*

Proof. We use the same notation with the proof of Theorem 3.1. By using (10), (11), and (12), we obtain

$$\begin{aligned} (\widetilde{\nabla} \log \pi_{\mathrm{SVS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{S}}(M)) &= \gamma(n-p-1)(\mathrm{tr}K)^{-1} I_p, \\ (\widetilde{\nabla} \log \pi_{\mathrm{S}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{S}}(M)) &= \gamma^2(\mathrm{tr}K)^{-2} K, \\ (\widetilde{\Delta} \log \pi_{\mathrm{S}}(M)) &= -n\gamma(\mathrm{tr}K)^{-1} I_p + 2\gamma(\mathrm{tr}K)^{-2} K. \end{aligned}$$

Therefore, from Lemma A.3,

$$\begin{aligned} &\mathrm{E}_M[(\hat{M}_{\mathrm{MSVS1}} - M)^\top(\hat{M}_{\mathrm{MSVS1}} - M)] - \mathrm{E}_M[(\hat{M}_{\mathrm{SVS}} - M)^\top(\hat{M}_{\mathrm{SVS}} - M)] \\ =&\frac{1}{N^2}(2\widetilde{\nabla} \log \pi_{\mathrm{SVS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{S}}(M) + \widetilde{\nabla} \log \pi_{\mathrm{S}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{S}}(M) + 2\widetilde{\Delta} \log \pi_{\mathrm{S}}(M)) \\ &\quad + o(N^{-2}) \\ =&\frac{1}{N^2}\gamma(\mathrm{tr}K)^{-2}(-2(p+1)(\mathrm{tr}K)I_p + (\gamma+4)K) + o(N^{-2}). \end{aligned}$$

Hence, we obtain (13). Since $K \preceq (\mathrm{tr}K)I_p$ from $K \succeq O$,

$$-2(p+1)(\mathrm{tr}K)I_p + (\gamma+4)K \preceq (\gamma - 2p + 2)(\mathrm{tr}K)I_p \prec O$$

if $0 < \gamma < 2p - 2$. $\square$

The generalized Bayes estimator with respect to π_{MSVS1} in (7) attains minimaxity in some cases as follows.

Theorem 3.3. *If $p \geq 2$, $p+2 \leq n < 2p+2-2/p$ and $0 < \gamma \leq -np+2p^2+2p-2$, then the generalized Bayes estimator with respect to π_{MSVS1} in (7) is minimax under the Frobenius loss.*

Proof. From Proposition B.1,

$$\Delta \pi_{\mathrm{MSVS1}}(M) = \gamma(\gamma + np - 2p^2 - 2p + 2) \|M\|_{\mathrm{F}}^{-2} \pi_{\mathrm{MSVS1}}(M) \leq 0.$$

Thus, $\pi_{\mathrm{MSVS1}}(M)$ is superharmonic, which indicates the minimaxity of the generalized Bayes estimator with respect to π_{MSVS1} in (7) under the Frobenius loss from Stein's classical result (Stein, 1974; Matsuda and Komaki, 2015). □

It is an interesting problem whether the generalized Bayes estimator with respect to π_{MSVS1} in (7) attains admissibility or not. In addition to Lemma 2.1, Brown (1971) derived the following sufficient condition for admissibility of generalized Bayes estimators, which may be useful here. While the condition (14) can be verified by using a similar argument to Theorem 2.1, the verification of the uniform boundedness of $\|\nabla \log m_\pi(y)\|$ seems difficult. We leave further investigation for future work.

Lemma 3.2. *(Brown, 1971) In estimation of θ from $Y \sim \mathrm{N}_d(\theta, I_d)$ under the quadratic loss, the generalized Bayes estimator of θ with respect to a prior $\pi(\theta)$ is admissible if $\|\nabla \log m_\pi(y)\|$ is uniformly bounded and*

$$\int_c^\infty \frac{r^{1-d}}{\overline{m}(r)} \mathrm{d}r = \infty, \tag{14}$$

where

$$\overline{m}(r) = \int m_\pi(y) \mathrm{d}U_r(y)$$

and U_r is the uniform measure on the sphere of radius r in $\mathbb{R}^d$.

4 Improvement by additional column-wise shrinkage

Here, instead of scalar shrinkage, we consider priors with additional column-wise shrinkage:

$$\pi_{\mathrm{MSVS2}}(M) = \pi_{\mathrm{SVS}}(M) \prod_{i=1}^p \|M_{\cdot i}\|^{-\gamma_i}, \tag{15}$$

where $\gamma_i \geq 0$ for every i and $\|M_{\cdot i}\|$ denotes the norm of the i-th column vector of M. Let

$$m_{\mathrm{MSVS2}}(Y) = \int p(Y \mid M) \pi_{\mathrm{MSVS2}}(M) \mathrm{d}M$$

be the marginal density of Y under the prior $\pi_{\mathrm{MSVS2}}(M)$.

Lemma 4.1. *If $0 \leq \gamma_i \leq p$ for every i, then $m_{\mathrm{MSVS2}}(Y) < \infty$ for every Y.*

Proof. Similarly to Lemma 3.1, it suffices to show that $\pi_{\mathrm{MSVS2}}(M)$ is locally integrable at $M = O$. Consider the neighborhood of $M = O$ defined by $A = \{M \mid \|M\|_{\mathrm{F}} \leq \varepsilon\}$ for $\varepsilon > 0$. To evaluate the integral on A, we use the variable transformation from M to $(r_1, \dots, r_p, u_1, \dots, u_p)$, where each $r_i \in [0, \infty)$ and $u_i \in \mathbb{R}^n$ with $\|u_i\| = 1$ are defined by $r_i = \|M_{\cdot i}\|$ and $u_i = M_{\cdot i}/r_i$ for the i-th column vector $M_{\cdot i}$ of M so that $M_{\cdot i} = r_i u_i$ (polar coordinate). Since $\mathrm{d}M_{\cdot i} = r_i^{n-1} \mathrm{d}r_i \mathrm{d}u_i$,

$$\mathrm{d}M = r_1^{n-1} \dots r_p^{n-1} \mathrm{d}r_1 \dots \mathrm{d}r_p \mathrm{d}u_1 \dots \mathrm{d}u_p.$$

Also, from $M^\top M = D(U^\top U)D$ with $D = \mathrm{diag}(r_1, \dots, r_p)$ and $U = (u_1 \dots u_p)$,

$$\begin{aligned}\pi_{\mathrm{MSVS2}}(M) &= \det(D)^{-(n-p-1)} \det(U^\top U)^{-(n-p-1)/2} \prod_{i=1}^p r_i^{-\gamma_i} \\ &= \prod_{i=1}^p r_i^{-(n-p-1)-\gamma_i} \cdot \det(U^\top U)^{-(n-p-1)/2}.\end{aligned}$$

Thus,

$$\begin{aligned}&\int_A \pi_{\mathrm{MSVS2}}(M) \mathrm{d}M \\ &= \int_{\|r\| \leq \varepsilon} \left(\prod_{i=1}^p r_i^{p-\gamma_i} \right) \mathrm{d}r_1 \dots \mathrm{d}r_p \int \det(U^\top U)^{-(n-p-1)/2} \mathrm{d}u_1 \dots \mathrm{d}u_p. \qquad (16)\end{aligned}$$

By variable transformation from $r = (r_1, \dots, r_p)$ to $s = \|r\|$ and $v = r/s$, the first integral in (16) is reduced to

$$\int_0^\varepsilon s^{p^2 - \sum_{i=1}^p \gamma_i + p - 1} \mathrm{d}s \int_{\|v\|=1} \left(\prod_{i=1}^p v_i^{p-\gamma_i} \right) \mathrm{d}v.$$

The integral with respect to s is finite if $p^2 - \sum_{i=1}^p \gamma_i + p - 1 > -1$, which is equivalent to $\sum_{i=1}^p \gamma_i < p^2 + p$. The integral with respect to v is finite if $p - \gamma_i \geq 0$ for every i. On the other hand, the second integral in (16) is finite due to the local integrability of π_{SVS}, which corresponds to $\gamma = 0$, at $M = O$. Therefore, $\pi_{\mathrm{MSVS2}}(M)$ is locally integrable at $M = O$ if $0 \leq \gamma_i \leq p$ for every i. □

From Lemma 4.1, the generalized Bayes estimator with respect to π_{MSVS2} is well-defined when $0 \leq \gamma \leq p$. We denote it by $\hat{M}_{\mathrm{MSVS2}}$.

Theorem 4.1. *For every M,*

$$N^2(\mathrm{E}_M[\|\hat{M}_{\mathrm{MSVS2}} - M\|_{\mathrm{F}}^2] - \mathrm{E}_M[\|\hat{M}_{\mathrm{SVS}} - M\|_{\mathrm{F}}^2]) \to \sum_{i=1}^p \gamma_i(\gamma_i - 2p + 2)\|M_{\cdot i}\|^{-2} \qquad (17)$$

as $N \to \infty$. Therefore, if $p \geq 2$ and $0 < \gamma_i \leq p$ for every i, then the generalized Bayes estimator with respect to π_{MSVS2} in (15) *asymptotically dominates that with respect to π_{SVS} in* (2) *under the Frobenius loss.*

Proof. Let

$$\pi_{\mathrm{CS}}(M) = \prod_{i=1}^{p} \|M_{\cdot i}\|^{-\gamma_i}.$$

Then,

$$\frac{\partial}{\partial M_{ai}} \log \pi_{\mathrm{CS}}(M) = -\gamma_i M_{ai} \|M_{\cdot i}\|^{-2}, \tag{18}$$

$$\frac{\partial^2}{\partial M_{ai}^2} \log \pi_{\mathrm{CS}}(M) = -\gamma_i \left(\|M_{\cdot i}\|^2 - 2M_{ai}^2 \right) \|M_{\cdot i}\|^{-4}. \tag{19}$$

From (10), (18), and (19),

$$\mathrm{tr}(\widetilde{\nabla} \log \pi_{\mathrm{SVS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M)) = (n-p-1) \sum_{i=1}^{p} \gamma_i \|M_{\cdot i}\|^{-2}, \tag{20}$$

$$\mathrm{tr}(\widetilde{\nabla} \log \pi_{\mathrm{CS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M)) = \sum_{i=1}^{p} \gamma_i^2 \|M_{\cdot i}\|^{-2}, \tag{21}$$

$$\mathrm{tr}(\widetilde{\Delta} \log \pi_{\mathrm{CS}}(M)) = -(n-2) \sum_{i=1}^{p} \gamma_i \|M_{\cdot i}\|^{-2}, \tag{22}$$

where we used the matrix derivative notations (28) and (29). Therefore, from Lemma A.2,

$$\begin{aligned}
&\mathrm{E}_M[\|\hat{M}_{\mathrm{MSVS2}} - M\|_{\mathrm{F}}^2] - \mathrm{E}_M[\|\hat{M}_{\mathrm{SVS}} - M\|_{\mathrm{F}}^2] \\
=&\frac{1}{N^2} \mathrm{tr}(2\widetilde{\nabla} \log \pi_{\mathrm{SVS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M) + \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M) + 2\widetilde{\Delta} \log \pi_{\mathrm{CS}}(M)) \\
&\quad + o(N^{-2}) \\
=&\frac{1}{N^2} \sum_{i=1}^{p} \gamma_i (\gamma_i - 2p + 2) \|M_{\cdot i}\|^{-2} + o(N^{-2}).
\end{aligned}$$

Hence, we obtain (17). □

From (17), the choice $\gamma_1 = \cdots = \gamma_p = p-1$ attains the minimum risk among $0 < \gamma_i \leq p$.

Now, we examine the performance of π_{MSVS2} in (15) by Monte Carlo simulation. Figure 4 plots the Frobenius risk of generalized Bayes estimators with respect to π_{MSVS2} in (15) with $\gamma_1 = \cdots = \gamma_p = p-1$, π_{MSVS1} in (7) with $\gamma = p^2 + p - 2$, π_{SVS} in (2) and $\pi_{\mathrm{S}}(M) = \|M\|_{\mathrm{F}}^{2-np}$, which is Stein's prior on the vectorization of M. We computed the generalized Bayes estimators by using the random walk Metropolis–Hastings algorithm with proposal variance 0.1. We set $M = U\Sigma$, where $U^\top U = I_p$ and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$. For this M, the Frobenius risk of the estimators compared here depends only on the singular values $\sigma_1, \ldots, \sigma_p$ of M. Overall, π_{MSVS2} performs better than π_{SVS}. Also, π_{MSVS2} even dominates π_{MSVS1} and π_{S} except when σ_1 is sufficiently small. This is understood from the column-wise shrinkage effect of π_{MSVS2}. Figure 5 plots the Frobenius risk for $n = 10$, $p = 3$ and $N = 10$, computed by the random walk Metropolis–Hastings algorithm with proposal variance 0.01. The risk behavior is similar to Figure 4. Figure 6 plots the Frobenius risk for $n = 20$, $p = 3$ and $N = 2$, computed by

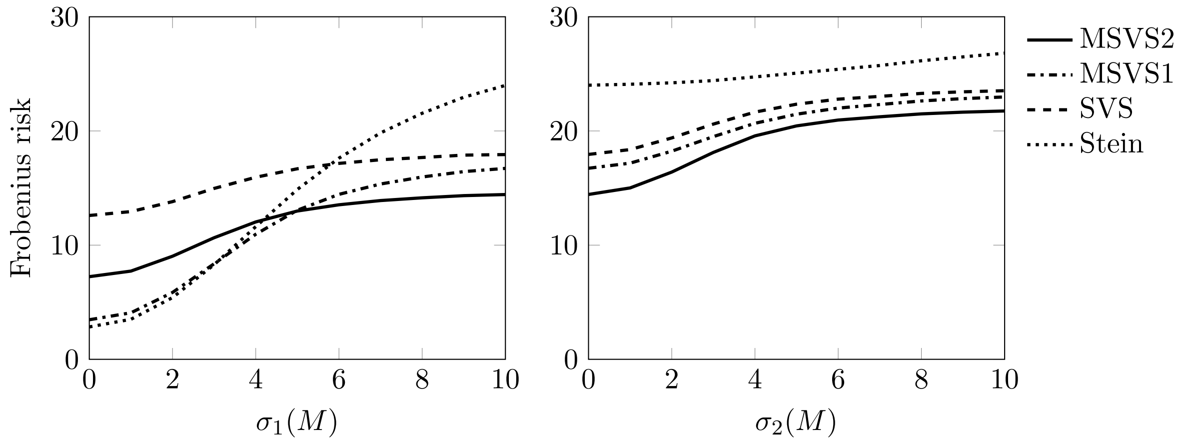

Figure 4: Frobenius risk of generalized Bayes estimators for $n = 10$, $p = 3$ and $N = 1$ where $M = U\Sigma$ with $U^\top U = I_p$ and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$. Left: $\sigma_2 = \sigma_3 = 0$. Right: $\sigma_1 = 10$, $\sigma_3 = 0$. solid: π_{MSVS2} with $\gamma_1 = \cdots = \gamma_p = p - 1$, dash-dotted: π_{MSVS1} with $\gamma = p^2 + p - 2$, dashed: π_{SVS}, dotted: Stein's prior π_{S}. Note that the minimax risk is $np/N = 30$.

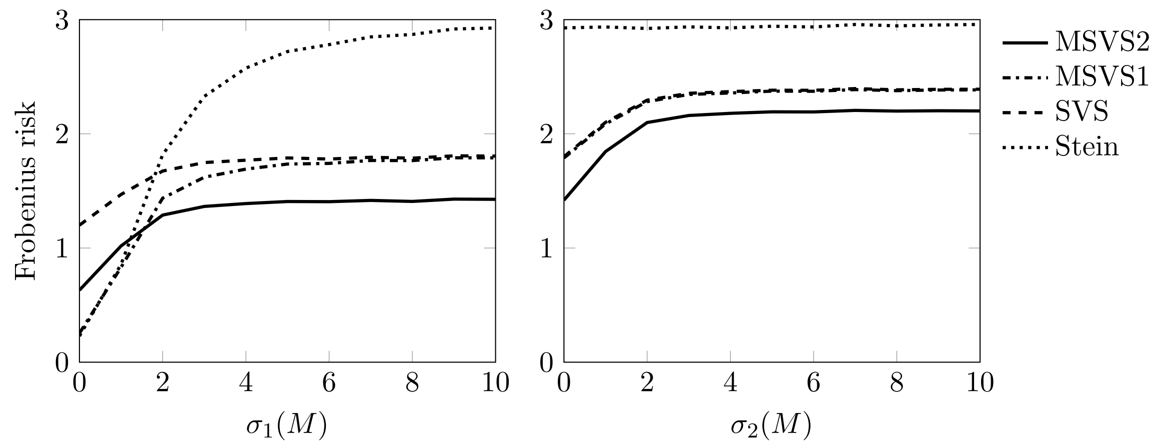

Figure 5: Frobenius risk of generalized Bayes estimators for $n = 10$, $p = 3$ and $N = 10$ where $M = U\Sigma$ with $U^\top U = I_p$ and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$. Left: $\sigma_2 = \sigma_3 = 0$. Right: $\sigma_1 = 10$, $\sigma_3 = 0$. solid: π_{MSVS2} with $\gamma_1 = \cdots = \gamma_p = p - 1$, dash-dotted: π_{MSVS1} with $\gamma = p^2 + p - 2$, dashed: π_{SVS}, dotted: Stein's prior π_{S}. Note that the minimax risk is $np/N = 3$.

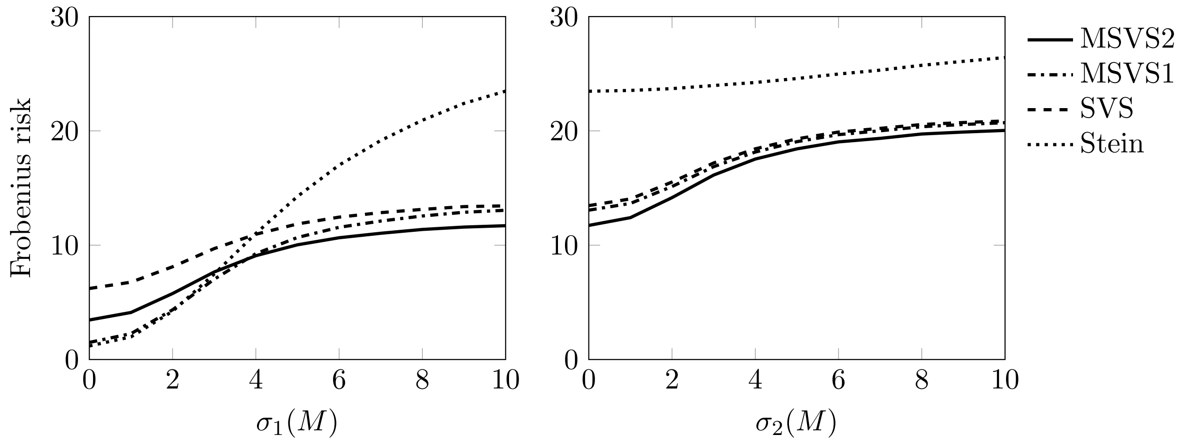

Figure 6: Frobenius risk of generalized Bayes estimators for $n = 20$, $p = 3$ and $N = 2$ where $M = U\Sigma$ with $U^\top U = I_p$ and $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_p)$. Left: $\sigma_2 = \sigma_3 = 0$. Right: $\sigma_1 = 10$, $\sigma_3 = 0$. solid: π_{MSVS2} with $\gamma_1 = \cdots = \gamma_p = p - 1$, dash-dotted: π_{MSVS1} with $\gamma = p^2 + p - 2$, dashed: π_{SVS}, dotted: Stein's prior π_{S}. Note that the minimax risk is $np/N = 30$.

the random walk Metropolis–Hastings algorithm with proposal variance 0.01. Again, the risk behavior is similar to Figure 4. Note that the value of $np/N = 30$ is the same with Figure 4.

Improvement by additional column-wise shrinkage holds even under the matrix quadratic loss (Matsuda and Strawderman, 2022; Matsuda, 2024).

Theorem 4.2. *For every M,*

$$\begin{aligned} N^2(\mathrm{E}_M[(\hat{M}_{\mathrm{MSVS2}} - M)^\top (\hat{M}_{\mathrm{MSVS2}} - M)] - \mathrm{E}_M[(\hat{M}_{\mathrm{SVS}} - M)^\top (\hat{M}_{\mathrm{SVS}} - M)]) \\ \to -2(p-1)D + DM^\top MD \end{aligned} \tag{23}$$

as $N \to \infty$, where $D = \mathrm{diag}(\gamma_1 \|M_{\cdot 1}\|^{-2}, \dots, \gamma_p \|M_{\cdot p}\|^{-2})$. Therefore, if $p \geq 2$ and $0 < \gamma_1 = \cdots = \gamma_p < 2 - 2/p$, then the generalized Bayes estimator with respect to π_{MSVS2} in (15) *asymptotically dominates that with respect to π_{SVS} in* (2) *under the matrix quadratic loss.*

Proof. We use the same notation with the proof of Theorem 4.1. By using (10), (18), and (19), we obtain

$$\begin{aligned} \widetilde{\nabla} \log \pi_{\mathrm{SVS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M) &= (n-p-1)D, \\ \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M) &= DM^\top MD, \\ \widetilde{\Delta} \log \pi_{\mathrm{CS}}(M) &= -(n-2)D. \end{aligned}$$

Therefore, from Lemma A.3,

$$\begin{aligned} &\mathrm{E}_M[(\hat{M}_{\mathrm{MSVS2}} - M)^\top (\hat{M}_{\mathrm{MSVS2}} - M)] - \mathrm{E}_M[(\hat{M}_{\mathrm{SVS}} - M)^\top (\hat{M}_{\mathrm{SVS}} - M)] \\ =& \frac{1}{N^2}(2\widetilde{\nabla} \log \pi_{\mathrm{SVS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M) + \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M)^\top \widetilde{\nabla} \log \pi_{\mathrm{CS}}(M) + 2\widetilde{\Delta} \log \pi_{\mathrm{CS}}(M)) \\ &+ o(N^{-2}) \\ =& -2(p-1)D + DM^\top MD. \end{aligned}$$

Hence, we obtain (23).

Suppose that $p \geq 2$ and $0 < \gamma_1 = \cdots = \gamma_p < 2 - 2/p$. Let $\| \cdot \|$ be the operater norm. Since $D \succeq O$ is diagonal and $\|M\|^2 \leq \|M\|_{\mathrm{F}}^2 = \sum_{i=1}^p \|M_{\cdot i}\|^2$,

$$
\begin{aligned}
\|D^{1/2} M^\top M D^{1/2}\| &\leq \|D^{1/2}\| \|M^\top M\| \|D^{1/2}\| \\
&= (\max_i D_{ii}) \|M\|^2 \\
&= \gamma_1 \frac{\|M\|^2}{\min_i \|M_{\cdot i}\|^2} \\
&\leq \gamma_1 \frac{\sum_{i=1}^p \|M_{\cdot i}\|^2}{\min_i \|M_{\cdot i}\|^2} \\
&\leq \gamma_1 p \\
&< 2(p-1),
\end{aligned}
$$

which yields $D^{1/2} M^\top M D^{1/2} \prec 2(p-1) I_p$. Therefore,

$$
-2(p-1)D + DM^\top MD = D^{1/2}(-2(p-1)I_p + D^{1/2} M^\top M D^{1/2}) D^{1/2} \prec O.
$$

□

The generalized Bayes estimator with respect to π_{MSVS2} in (15) attains minimaxity in some cases as follows. It is an interesting future work to investigate its admissibility.

Theorem 4.3. *If $p \geq 3$, $p + 2 \leq n < 2p$ and $0 < \gamma \leq -n + 2p$, then the generalized Bayes estimator with respect to π_{MSVS2} in (15) is minimax under the Frobenius loss.*

Proof. From Proposition B.2,

$$
\Delta \pi_{\mathrm{MSVS2}}(M) = \gamma(\gamma + n - 2p) \left(\sum_{i=1}^p \|M_{\cdot i}\|^{-2} \right) \pi_{\mathrm{MSVS2}}(M) \leq 0.
$$

Thus, $\pi_{\mathrm{MSVS2}}(M)$ is superharmonic, which indicates the minimaxity of the generalized Bayes estimator with respect to π_{MSVS2} in (15) under the Frobenius loss from Stein's classical result (Stein, 1974; Matsuda and Komaki, 2015). □

5 Bayesian prediction

Here, we consider Bayesian prediction and provide parallel results to those in Sections 3 and 4. Suppose that we observe $Y \sim \mathrm{N}_{n,p}(M, I_n, N^{-1} I_p)$ and predict $\widetilde{Y} \sim \mathrm{N}_{n,p}(M, I_n, I_p)$ by a predictive density $\hat{p}(\widetilde{Y} \mid Y)$. We evaluate predictive densities by the Kullback–Leibler loss:

$$
D(p(\cdot \mid M), \hat{p}(\cdot \mid Y)) = \int p(\widetilde{Y} \mid M) \log \frac{p(\widetilde{Y} \mid M)}{\hat{p}(\widetilde{Y} \mid Y)} \mathrm{d}\widetilde{Y}.
$$

The Bayesian predictive density based on a prior $\pi(M)$ is defined as

$$
\hat{p}_\pi(\widetilde{Y} \mid Y) = \int p(\widetilde{Y} \mid M) \pi(M \mid Y) \mathrm{d}M,
$$

where $\pi(M \mid Y)$ is the posterior distribution of M given Y, and it minimizes the Bayes risk (Aitchison, 1975):

$$\hat{p}_\pi(\widetilde{Y} \mid Y) = \arg\min_{\hat{p}} \int D(p(\cdot \mid M), \hat{p}(\cdot \mid Y)) p(Y \mid M)\pi(M) \mathrm{d}Y \mathrm{d}M.$$

The Bayesian predictive density with respect to the uniform prior is minimax. However, it is inadmissible and dominated by Bayesian predictive densities based on superharmonic priors (Komaki, 2001; George, Liang and Xu, 2006). In particular, the Bayesian predictive density based on the singular value shrinkage prior π_{SVS} in (2) is minimax and dominates that based on the uniform prior (Matsuda and Komaki, 2015).

The asymptotic expansion of the difference between the Kullback–Leibler risk of two Bayesian predictive densities is obtained as follows.

Lemma 5.1. *As $N \to \infty$, the difference between the Kullback–Leibler risk of $p_{\pi_1}(\widetilde{Y} \mid Y)$ and $p_{\pi_1\pi_2}(\widetilde{Y} \mid Y)$ is expanded as*

$$\begin{aligned}
&\mathrm{E}_M[D(p(\widetilde{Y} \mid M), p_{\pi_1\pi_2}(\widetilde{Y} \mid Y))] - \mathrm{E}_M[D(p(\widetilde{Y} \mid M), p_{\pi_1}(\widetilde{Y} \mid Y))] \\
=&\frac{1}{2N^2}\mathrm{tr}(2(\widetilde{\nabla} \log \pi_1(M))^\top (\widetilde{\nabla} \log \pi_2(M)) + (\widetilde{\nabla} \log \pi_2(M))^\top (\widetilde{\nabla} \log \pi_2(M)) + 2\widetilde{\Delta} \log \pi_2(M)) \\
&+ o(N^{-2}).
\end{aligned} \tag{24}$$

Proof. For the normal model with known covariance, the information geometrical quantities (Amari, 1985) are given by

$$g_{ij} = g^{ij} = \delta_{ij}, \quad \Gamma_{ij}^k = 0, \quad T_{ijk} = 0.$$

Also, the Jeffreys prior coincides with the uniform prior $\pi(M) \equiv 1$. Therefore, from equation (3) of Komaki (2006), the Kullback–Leibler risk of the Bayesian predictive density $p_\pi(\widetilde{Y} \mid Y)$ based on a prior $\pi(M)$ is expanded as

$$\begin{aligned}
&\mathrm{E}_M[D(p(\widetilde{Y} \mid M), p_\pi(\widetilde{Y} \mid Y))] \\
=&\frac{np}{2N} + \frac{1}{2N^2}\mathrm{tr}((\widetilde{\nabla} \log \pi(M))^\top (\widetilde{\nabla} \log \pi(M)) + 2\widetilde{\Delta} \log \pi(M)) + g(M) + o(N^{-2}),
\end{aligned} \tag{25}$$

where $g(M)$ is a function independent of $\pi(M)$. Substituting $\pi = \pi_1\pi_2$ and $\pi = \pi_1$ into (25) and taking difference, we obtain (24). □

By comparing Lemma 5.1 to Lemma A.2, we obtain the following connection between estimation and prediction.

Proposition 5.1. *For every M,*

$$\begin{aligned}
&\lim_{N\to\infty} N^2(\mathrm{E}_M[D(p(\widetilde{Y} \mid M), \hat{p}_{\pi_1\pi_2}(\widetilde{Y} \mid Y))] - \mathrm{E}_M[D(p(\widetilde{Y} \mid M), \hat{p}_{\pi_1}(\widetilde{Y} \mid Y))]) \\
=&\frac{1}{2}\lim_{N\to\infty} N^2(\mathrm{E}_M[\|\hat{M}^{\pi_1\pi_2} - M\|_{\mathrm{F}}^2] - \mathrm{E}_M[\|\hat{M}^{\pi_1} - M\|_{\mathrm{F}}^2]).
\end{aligned}$$

Therefore, if $\hat{\theta}^{\pi_1\pi_2}$ asymptotically dominates $\hat{\theta}^{\pi_1}$ under the quadratic loss, then $\hat{p}_{\pi_1\pi_2}(\widetilde{Y} \mid Y)$ asymptotically dominates $\hat{p}_{\pi_1}(\widetilde{Y} \mid Y)$ under the Kullback–Leibler loss.

Therefore, Theorems 3.1 and 4.1 are extended to Bayesian prediction as follows. Other results in the previous sections can be extended to Bayesian prediction similarly.

Theorem 5.1. *For every M,*

$$\begin{aligned}&N^2(\mathrm{E}_M[D(p(\cdot \mid M), \hat{p}_{\mathrm{MSVS1}}(\cdot \mid Y))] - \mathrm{E}_M[D(p(\cdot \mid M), \hat{p}_{\mathrm{SVS}}(\cdot \mid Y))])\\ \to&\frac{\gamma(\gamma - 2p^2 - 2p + 4)}{2\mathrm{tr}(M^\top M)}\end{aligned}$$

as $N \to \infty$. Therefore, if $p \geq 2$ and $0 < \gamma < p^2 + p$, then the Bayesian predictive density with respect to π_{MSVS1} in (7) *asymptotically dominates that with respect to π_{SVS} in* (2) *under the Kullback–Leibler loss.*

Theorem 5.2. *For every M,*

$$\begin{aligned}&N^2(\mathrm{E}_M[D(p(\cdot \mid M), \hat{p}_{\mathrm{MSVS2}}(\cdot \mid Y))] - \mathrm{E}_M[D(p(\cdot \mid M), \hat{p}_{\mathrm{SVS}}(\cdot \mid Y))])\\ \to&\frac{1}{2}\sum_{i=1}^p \gamma_i(\gamma_i - 2p + 2)\|M_{\cdot i}\|^{-2}\end{aligned}$$

as $N \to \infty$. Therefore, if $p \geq 2$ and $0 < \gamma \leq p$, then the Bayesian predictive density with respect to π_{MSVS2} in (15) *asymptotically dominates that with respect to π_{SVS} in* (2) *under the Kullback–Leibler loss.*

Figures 7 and 8 plot the Kullback–Leibler risk of Bayesian predictive densities in similar settings to Figures 1 and 4, respectively. They show that the risk behavior in prediction is qualitatively the same with that in estimation, which is compatible with Theorems 5.1 and 5.2.

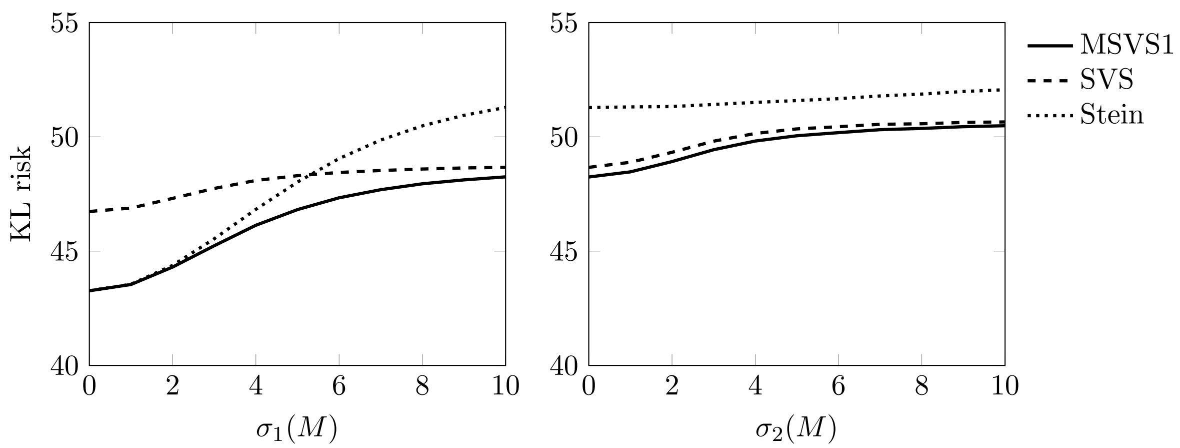

Figure 7: Kullback–Leibler risk of Bayesian predictive densities for $n = 10$, $p = 3$ and $N = 1$. Left: $\sigma_2 = \sigma_3 = 0$. Right: $\sigma_1 = 10$, $\sigma_3 = 0$. solid: π_{MSVS1} with $\gamma = p^2 + p - 2$, dashed: π_{SVS}, dotted: Stein's prior π_{S}.

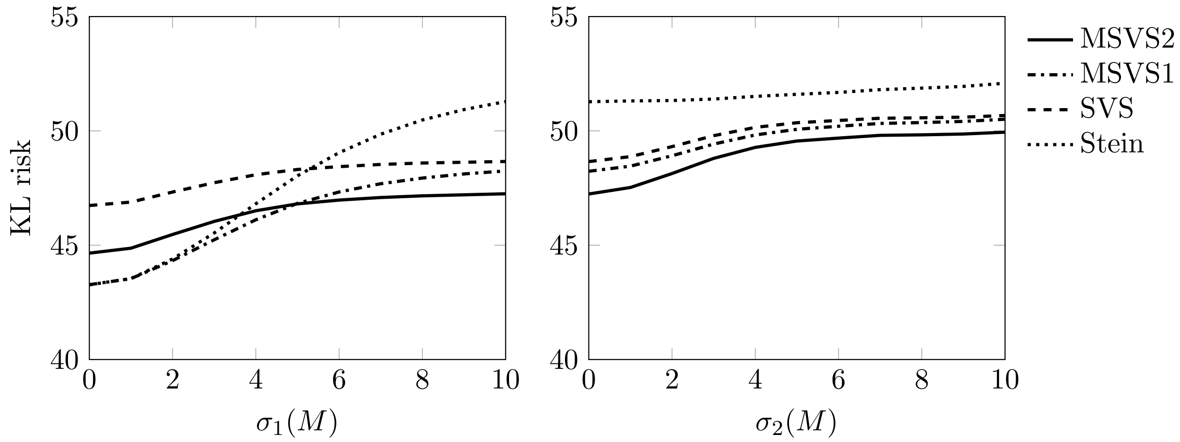

Figure 8: Kullback–Leibler risk of Bayesian predictive densities for $n = 10$, $p = 3$ and $N = 1$ where $M = U\Sigma$ with $U^\top U = I_p$ and $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_p)$. Left: $\sigma_2 = \sigma_3 = 0$. Right: $\sigma_1 = 10$, $\sigma_3 = 0$. solid: π_{MSVS2} with $\gamma_1 = \cdots = \gamma_p = p - 1$, dash-dotted: π_{MSVS1} with $\gamma = p^2 + p - 2$, dashed: π_{SVS}, dotted: Stein's prior π_{S}.

Acknowledgments

We thank the referees for helpful comments. This work was supported by JSPS KAKENHI Grant Numbers 19K20220, 21H05205, 22K17865 and JST Moonshot Grant Number JPMJMS2024.

References

Aitchison, J. (1975). Goodness of prediction fit. *Biometrika* **62**, 547–554.

Amari, S. (1985). *Differential-Geometrical Methods in Statistics.* New York: Springer.

Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Annals of Mathematical Statistics* **36**, 855–903.

Brown, L. D. & Zhao, L. H. (2009). Estimators for Gaussian models having a block-wise structure. *Statistica Sinica* **19**, 885–903.

Efron, B. & Morris, C. (1972). Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika* **59**, 335–347.

Efron, B. & Morris, C. (1976). Multivariate empirical Bayes and estimation of covariance matrices. *Annals of Statistics* **4**, 22–32.

Fourdrinier, D., Strawderman, W. E. & Wells, M. (2018). *Shrinkage Estimation.* Springer, New York.

George, E. I., Liang, F. & Xu, X. (2006). Improved minimax predictive densities under Kullback–Leibler loss. *Annals of Statistics* **34**, 78–91.

Gupta, A. K. & Nagar, D. K. (2000). *Matrix Variate Distributions*. New York: Chapman & Hall.

Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika* **88**, 859–864.

Komaki, F. (2006). Shrinkage priors for Bayesian prediction. *Annals of Statistics* **34**, 808–819.

Konno, Y. (1990). Families of minimax estimators of matrix of normal means with unknown covariance matrix. *J. Japan Statist. Soc.* **20**, 191–201.

Konno, Y. (1991). On estimation of a matrix of normal means with unknown covariance matrix. *Journal of Multivariate Analysis* **36**, 44–55.

Matsuda, T. (2023). Adapting to arbitrary quadratic loss via singular value shrinkage. *IEEE Transactions on Information Theory*, accepted.

Matsuda, T. (2024). Matrix quadratic risk of orthogonally invariant estimators for a normal mean matrix. *Japanese Journal of Statistics and Data Science*, accepted.

Matsuda, T. & Komaki, F. (2015). Singular value shrinkage priors for Bayesian prediction. *Biometrika* **102**, 843–854.

Matsuda, T. & Strawderman, W. E. (2019). Improved loss estimation for a normal mean matrix. *Journal of Multivariate Analysis* **169**, 300–311.

Matsuda, T. & Strawderman, W. E. (2022). Estimation under matrix quadratic loss and matrix superharmonicity. *Biometrika* **109**, 503–519.

Stein, C. (1974). Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symp. Asymptotic Statistics* **2**, 345–381.

Tsukuma, H. & Kubokawa, T. (2007). Methods for improvement in estimation of a normal mean matrix. *Journal of Multivariate Analysis* **98**, 1592–1610.

Tsukuma, H. & Kubokawa, T. (2017). Proper Bayes and minimax predictive densities related to estimation of a normal mean matrix. *Journal of Multivariate Analysis* **159**, 138–150.

Tsukuma, H. & Kubokawa, T. (2020). *Shrinkage estimation for mean and covariance matrices*. Springer.

Yuasa, R. & Kubokawa, T. (2023). Generalized Bayes estimators with closed forms for the normal mean and covariance matrices. *Journal of Statistical Planning and Inference* **222**, 182–194.

Yuasa, R. & Kubokawa, T. (2023). Weighted shrinkage estimators of normal mean matrices and dominance properties. *Journal of Multivariate Analysis* **194**, 1–17.

A Asymptotic expansion of risk

Here, we provide asymptotic expansion formulas for estimators of a normal mean vector. Consider the problem of estimating θ from the observation $Y \sim \mathrm{N}_d(\theta, N^{-1} I_d)$ under the quadratic loss $l(\theta, \hat{\theta}) = \|\hat{\theta} - \theta\|^2$. As shown in Stein (1974), the generalized Bayes estimator $\hat{\theta}^{\pi}$ with respect to a prior $\pi(\theta)$ is expressed as

$$\hat{\theta}^{\pi}(y) = y + \frac{1}{N} \nabla_y \log m_{\pi}(y),$$

where

$$m_{\pi}(y) = \int p(y \mid \theta) \pi(\theta) \mathrm{d}\theta.$$

The asymptotic difference between the quadratic risk of two generalized Bayes estimators as $N \to \infty$ is given as follows.

Lemma A.1. *As $N \to \infty$, the difference between the quadratic risk of $\hat{\theta}^{\pi_1}$ and $\hat{\theta}^{\pi_1 \pi_2}$ is expanded as*

$$\begin{aligned} & \mathrm{E}_{\theta}[\|\hat{\theta}^{\pi_1 \pi_2} - \theta\|^2] - \mathrm{E}_{\theta}[\|\hat{\theta}^{\pi_1} - \theta\|^2] \\ =& \frac{1}{N^2} (2 \nabla \log \pi_1(\theta)^{\top} \nabla \log \pi_2(\theta) + \|\nabla \log \pi_2(\theta)\|^2 + 2\Delta \log \pi_2(\theta)) + o(N^{-2}). \end{aligned} \tag{26}$$

Proof. By using Stein's lemma (Fourdrinier et al., 2018) and $m_{\pi}(y) = \pi(y) + o(1)$ as $N \to \infty$, the quadratic risk of the generalized Bayes estimator $\hat{\theta}^{\pi}$ is calculated as

$$\begin{aligned} & \mathrm{E}_{\theta}[\|\hat{\theta}^{\pi}(y) - \theta\|^2] \\ =& \mathrm{E}_{\theta}[\|y - \theta\|^2] + \frac{2}{N} \mathrm{E}_{\theta}[(y - \theta)^{\top} \nabla \log m_{\pi}(y)] + \frac{1}{N^2} \mathrm{E}_{\theta}[\|\nabla \log m_{\pi}(y)\|^2] \\ =& \frac{d}{N} + \frac{1}{N^2} \mathrm{E}_{\theta}[\|\nabla \log m_{\pi}(y)\|^2 + 2\Delta \log m_{\pi}(y)] \\ =& \frac{d}{N} + \frac{1}{N^2} \left(\|\nabla \log \pi(\theta)\|^2 + 2\Delta \log \pi(\theta) \right) + o(N^{-2}). \end{aligned} \tag{27}$$

Substituting $\pi = \pi_1 \pi_2$ and $\pi = \pi_1$ into (27) and taking difference, we obtain (26). □

We extend the above formula to matrices by using the matrix derivative notations from Matsuda and Strawderman (2022). For a function $f : \mathbb{R}^{n \times p} \to \mathbb{R}$, its matrix gradient $\widetilde{\nabla} f : \mathbb{R}^{n \times p} \to \mathbb{R}^{n \times p}$ is defined as

$$(\widetilde{\nabla} f(X))_{ai} = \frac{\partial}{\partial X_{ai}} f(X). \tag{28}$$

For a C^2 function $f : \mathbb{R}^{n \times p} \to \mathbb{R}$, its matrix Laplacian $\widetilde{\Delta} f : \mathbb{R}^{n \times p} \to \mathbb{R}^{p \times p}$ is defined as

$$(\widetilde{\Delta} f(X))_{ij} = \sum_{a=1}^{n} \frac{\partial^2}{\partial X_{ai} \partial X_{aj}} f(X). \tag{29}$$

Then, the above formulas can be straightforwardly extended to matrix-variate normal distributions as follows.

Lemma A.2. *As $N \to \infty$, the difference between the Frobenius risk of $\hat{M}^{\pi_1}$ and $\hat{M}^{\pi_1 \pi_2}$ is expanded as*

$$
\begin{aligned}
&\mathrm{E}_M[\|\hat{M}^{\pi_1\pi_2} - M\|_{\mathrm{F}}^2] - \mathrm{E}_M[\|\hat{M}^{\pi_1} - M\|_{\mathrm{F}}^2] \\
=&\frac{1}{N^2}\mathrm{tr}(2\widetilde{\nabla}\log\pi_1(M)^\top\widetilde{\nabla}\log\pi_2(M) + \widetilde{\nabla}\log\pi_2(M)^\top\widetilde{\nabla}\log\pi_2(M) + 2\widetilde{\Delta}\log\pi_2(M)) \\
&+ o(N^{-2}).
\end{aligned}
$$

Lemma A.3. *As $N \to \infty$, the difference between the matrix quadratic risk of $\hat{M}^{\pi_1}$ and $\hat{M}^{\pi_1\pi_2}$ is expanded as*

$$
\begin{aligned}
&\mathrm{E}_M[(\hat{M}^{\pi_1\pi_2} - M)^\top(\hat{M}^{\pi_1\pi_2} - M) - (\hat{M}^{\pi_1} - M)^\top(\hat{M}^{\pi_1} - M)] \\
=&\frac{1}{N^2}(2\widetilde{\nabla}\log\pi_1(M)^\top\widetilde{\nabla}\log\pi_2(M) + \widetilde{\nabla}\log\pi_2(M)^\top\widetilde{\nabla}\log\pi_2(M) + 2\widetilde{\Delta}\log\pi_2(M)) \\
&+ o(N^{-2}).
\end{aligned}
$$

Komaki (2006) derived the asymptotic expansion of the Kullback–Leibler risk of Bayesian predictive densities. For the normal model as discussed in Section 5, the result shows that Stein's prior dominates the Jeffreys prior in $O(N^{-1})$ term at the origin and $O(N^{-2})$ term at other points, which is reminiscent of superefficiency theory. A similar phenomenon should exist in estimation as well. Unlike Stein's prior, the priors for a normal mean matrix such as π_{SVS} diverge at many points such as low-rank matrices. It is an interesting future problem to investigate the asymptotic risk of such priors in detail.

B Laplacian of π_{MSVS1} and π_{MSVS2}

Lemma B.1. *(Stein, 1974; Matsuda and Strawderman, 2019) Suppose that $f : \mathbb{R}^{n\times p} \to \mathbb{R}$ is represented as $f(X) = \widetilde{f}(\sigma)$, where $n \geq p$ and $\sigma = (\sigma_1(X), \ldots, \sigma_p(X))$ denotes the singular values of X. If f is twice weakly differentiable, then its Laplacian is*

$$
\Delta f = \sum_{a=1}^n \sum_{i=1}^p \frac{\partial^2 f}{\partial X_{ai}^2} = 2\sum_{i<j} \frac{\sigma_i \partial\widetilde{f}/\partial\sigma_i - \sigma_j\partial\widetilde{f}/\partial\sigma_j}{\sigma_i^2 - \sigma_j^2} + (n-p)\sum_{i=1}^p \frac{1}{\sigma_i}\frac{\partial\widetilde{f}}{\partial\sigma_i} + \sum_{i=1}^p \frac{\partial^2\widetilde{f}}{\partial\sigma_i^2}.
$$

Proposition B.1. *The Laplacian of π_{MSVS1} in (7) is given by*

$$
\Delta\pi_{\mathrm{MSVS1}}(M) = \gamma(\gamma + np - 2p^2 - 2p + 2)\|M\|_{\mathrm{F}}^{-2}\pi_{\mathrm{MSVS1}}(M).
$$

Proof. Let

$$
\widetilde{f}(\sigma) = \left(\prod_{i=1}^p \sigma_i^{-(n-p-1)}\right)\left(\sum_{i=1}^p \sigma_i^2\right)^{-\gamma/2}
$$

so that $\pi_{\mathrm{MSVS1}}(M) = \widetilde{f}(\sigma)$ with $\sigma = (\sigma_1(M), \ldots, \sigma_p(M))$. From Lemma B.1 and

$$
\frac{\partial\widetilde{f}}{\partial\sigma_i} = -(n-p-1)\sigma_i^{-1}\widetilde{f} - \gamma\sigma_i\left(\sum_{j=1}^p \sigma_j^2\right)^{-1}\widetilde{f},
$$

$$\frac{\partial^2 \widetilde{f}}{\partial \sigma_i^2} = (n-p)(n-p-1)\sigma_i^{-2}\widetilde{f} + (2n-2p-3)\gamma \left(\sum_{j=1}^p \sigma_j^2 \right)^{-1} \widetilde{f} + \gamma(\gamma+2)\sigma_i^2 \left(\sum_{j=1}^p \sigma_j^2 \right)^{-2} \widetilde{f},$$

we have

$$\begin{aligned}
\Delta \pi_{\mathrm{MSVS1}}(M) &= 2 \sum_{i<j} \frac{\sigma_i \partial \widetilde{f}/\partial \sigma_i - \sigma_j \partial \widetilde{f}/\partial \sigma_j}{\sigma_i^2 - \sigma_j^2} + (n-p) \sum_{i=1}^p \frac{1}{\sigma_i} \frac{\partial \widetilde{f}}{\partial \sigma_i} + \sum_{i=1}^p \frac{\partial^2 \widetilde{f}}{\partial \sigma_i^2} \\
&= -2 \cdot \frac{p(p-1)}{2} \gamma \left(\sum_{i=1}^p \sigma_i^2 \right)^{-1} \widetilde{f} - (n-p)(n-p-1) \left(\sum_{i=1}^p \sigma_i^{-2} \right) \widetilde{f} \\
&\quad - p(n-p)\gamma \left(\sum_{i=1}^p \sigma_i^2 \right)^{-1} \widetilde{f} + (n-p)(n-p-1) \left(\sum_{i=1}^p \sigma_i^{-2} \right) \widetilde{f} \\
&\quad + \gamma(p(2n-2p-3) + \gamma + 2) \left(\sum_{i=1}^p \sigma_i^2 \right)^{-1} \widetilde{f} \\
&= \gamma(\gamma + np - 2p^2 - 2p + 2) \left(\sum_{i=1}^p \sigma_i^2 \right)^{-1} \widetilde{f}.
\end{aligned}$$

□

Proposition B.2. *The Laplacian of* π_{MSVS2} *in* (15) *is given by*

$$\Delta \pi_{\mathrm{MSVS2}}(M) = \gamma(\gamma + n - 2p) \left(\sum_{i=1}^p \| M_{\cdot i} \|^{-2} \right) \pi_{\mathrm{MSVS2}}(M).$$

Proof. From

$$\Delta \log f = \frac{\Delta f}{f} - \| \nabla \log f \|^2$$

and (20), (21) and (22),

$$\begin{aligned}
\frac{\Delta \pi_{\mathrm{MSVS2}}(M)}{\pi_{\mathrm{MSVS2}}(M)} &= \Delta \log \pi_{\mathrm{MSVS2}}(M) + \| \nabla \log \pi_{\mathrm{MSVS2}}(M) \|^2 \\
&= \Delta \log \pi_{\mathrm{SVS}}(M) + \Delta \log \pi_{\mathrm{CS}}(M) + \| \nabla \log \pi_{\mathrm{SVS}}(M) + \nabla \log \pi_{\mathrm{CS}}(M) \|^2 \\
&= \frac{\Delta \pi_{\mathrm{SVS}}(M)}{\pi_{\mathrm{SVS}}(M)} + (-(n-2)\gamma + 2(n-p-1)\gamma + \gamma^2) \sum_{i=1}^p \| M_{\cdot i} \|^{-2} \\
&= \gamma(\gamma + n - 2p) \sum_{i=1}^p \| M_{\cdot i} \|^{-2},
\end{aligned}$$

where we used $\Delta \pi_{\mathrm{SVS}}(M) = 0$ (Theorem 2 of Matsuda and Komaki, 2015). □

C Improving on the block-wise Stein prior

Here, we develop priors that asymptotically dominate the block-wise Stein prior in estimation and prediction. Suppose that we observe $Y \sim \mathrm{N}_d(\theta, N^{-1} I_d)$ and estimate θ or predict $\widetilde{Y} \sim \mathrm{N}_d(\theta, I_d)$. We assume that the d-dimensional mean vector θ is split into B disjoint blocks $\theta^{(1)}, \ldots, \theta^{(B)}$ with size $d_1, \ldots, d_B$, where $d_1 + \cdots + d_B = d$. For example. such a situation appears in balanced ANOVA and wavelet regression (Brown and Zhao, 2009). Then, the block-wise Stein prior is defined as

$$\pi_{\mathrm{BS}}(\theta) = \prod_{b=1}^{B} \|\theta^{(b)}\|^{R_b}, \quad R_b = -(d_b - 2)_+, \tag{30}$$

which puts Stein's prior on each block. Since it is superharmonic, the generalized Bayes estimator $\hat{\theta}^{\pi_{\mathrm{BS}}}$ with respect to π_{BS} is minimax. However, Brown and Zhao (2009) showed that $\hat{\theta}^{\pi_{\mathrm{BS}}}$ is inadmissible and dominated by an estimator with additional James–Stein type shrinkage defined by

$$\hat{\theta}(y) = \hat{\theta}^{\pi_{\mathrm{BS}}}(y) - \frac{R_\# + d - 2}{\|y\|^2} y,$$

where $R_\# = \sum_b R_b > 2 - d$. From this result, Brown and Zhao (2009) conjectured that the block-wise Stein prior can be improved by multiplying a Stein-type shrinkage prior in Remark 3.2. Following their conjecture, we construct priors by adding scalar shrinkage to the block-wise Stein priors:

$$\pi_{\mathrm{MBS}}(\theta) = \pi_{\mathrm{BS}}(\theta) \|\theta\|^{-\gamma}, \tag{31}$$

where $\gamma \geq 0$. Let

$$m_{\mathrm{MBS}}(y) = \int p(y \mid \theta) \pi_{\mathrm{MBS}}(\theta) \mathrm{d}\theta.$$

Lemma C.1. *If $0 \leq \gamma < B(R_b + d_b)$ for every b, then $m_{\mathrm{MBS}}(y) < \infty$ for every y.*

Proof. Since $m_{\mathrm{MBS}}(y)$ is interpreted as the expectation of $\pi_{\mathrm{MBS}}(\theta)$ under $\theta \sim \mathrm{N}_d(y, I_d)$, it suffices to show that $\pi_{\mathrm{MBS}}(\theta)$ is locally integrable at every θ.

First, consider $\theta \neq 0$. Since $m_{\mathrm{BS}}(y) < \infty$ for every y (Brown and Zhao, 2009), $\pi_{\mathrm{BS}}(\theta)$ is locally integrable at θ. Also, $\|\theta\| > c$ for some $c > 0$ in a neighborhood of θ. Thus, $\pi_{\mathrm{MBS}}(\theta) = \pi_{\mathrm{BS}}(\theta) \|\theta\|^{-\gamma}$ is locally integrable at θ.

Next, consider $\theta = 0$ and take the neighborhood $A = \{\theta \mid \|\theta^{(1)}\| \leq s, \ldots, \|\theta^{(B)}\| \leq s\}$ for $s > 0$. From the AM-GM inequality,

$$\|\theta\|^2 = \sum_b \|\theta^{(b)}\|^2 \geq B \left(\prod_b \|\theta^{(b)}\|^2 \right)^{1/B} = B \prod_b \|\theta^{(b)}\|^{2/B}.$$

Thus,

$$\int_A \pi_{\mathrm{MBS}}(\theta) \mathrm{d}\theta \leq C \int_0^s \cdots \int_0^s \prod_b r_b^{R_b + d_b - 1 - \gamma/B} \mathrm{d}r_1 \cdots \mathrm{d}r_B,$$

where $r_b = \|\theta^{(b)}\|$ and C is a constant. Therefore, $\pi_{\mathrm{MBS}}(\theta)$ is locally integrable at $\theta = 0$ if $R_b + d_b - 1 - \gamma/B > -1$ for every b, which is equivalent to $\gamma < B(R_b + d_b)$ for every b. $\square$

From Lemma C.1, the generalized Bayes estimator with respect to π_{MBS} is well-defined when $0 \leq \gamma < B(R_b + d_b)$ for every b. We denote it by $\hat{\theta}_{\mathrm{MBS}}$.

Theorem C.1. *For every M,*

$$N^2(\mathrm{E}_\theta[\|\hat{\theta}_{\mathrm{MBS}} - \theta\|^2] - \mathrm{E}_\theta[\|\hat{\theta}_{\mathrm{BS}} - \theta\|^2]) \to \gamma(\gamma - 2(R_\# + d - 2))\|\theta\|^{-2} \tag{32}$$

as $N \to \infty$. Therefore, if $0 < \gamma < 2(R_\# + d - 2)$, then the generalized Bayes estimator with respect to π_{MBS} in (31) *asymptotically dominates that with respect to π_{BS} in* (30) *under the Frobenius loss.*

Proof. Let $\pi_{\mathrm{S}}(\theta) = \|\theta\|^{-\gamma}$. By straightforward calculation, we obtain

$$\begin{aligned}\nabla \log \pi_{\mathrm{BS}}(\theta)^\top \nabla \log \pi_{\mathrm{S}}(\theta) &= -\gamma R_\# \|\theta\|^{-2},\\ \nabla \log \pi_{\mathrm{S}}(\theta)^\top \nabla \log \pi_{\mathrm{S}}(\theta) &= \gamma^2 \|\theta\|^{-2},\\ \Delta \log \pi_{\mathrm{S}}(\theta) &= -\gamma(d-2)\|\theta\|^{-2}.\end{aligned}$$

Therefore, from Lemma A.1,

$$\begin{aligned}&\mathrm{E}_\theta[\|\hat{\theta}_{\mathrm{MBS}} - \theta\|^2] - \mathrm{E}_\theta[\|\hat{\theta}_{\mathrm{BS}} - \theta\|^2]\\ =&\frac{1}{N^2}\left(2\nabla \log \pi_{\mathrm{BS}}(\theta)^\top \nabla \log \pi_{\mathrm{S}}(\theta) + \|\nabla \log \pi_{\mathrm{S}}(\theta)\|^2 + 2\Delta \log \pi_{\mathrm{S}}(\theta)\right) + o(N^{-2})\\ =&\frac{1}{N^2}\gamma(\gamma - 2(R_\# + d - 2))\|\theta\|^{-2} + o(N^{-2}).\end{aligned}$$

Hence, we obtain (32). □

From (32), the choice $\gamma = R_\# + d - 2$ is optimal. As discussed in Section 5, Theorem C.1 is extended to Bayesian prediction as follows.

Theorem C.2. *For every M,*

$$N^2(\mathrm{E}_\theta[D(p(\cdot \mid \theta), \hat{p}_{\mathrm{MBS}}(\cdot \mid y))] - \mathrm{E}_\theta[D(p(\cdot \mid \theta), \hat{p}_{\mathrm{BS}}(\cdot \mid y))]) \to \frac{\gamma(\gamma - 2(R_\# + d - 2))}{2}\|\theta\|^{-2}$$

as $N \to \infty$. Therefore, if $p \geq 2$ and $0 < \gamma < p^2 + p$, then the Bayesian predictive density with respect to π_{MBS} in (31) *asymptotically dominates that with respect to π_{BS} in* (30) *under the Kullback–Leibler loss.*