Shrinkage Estimation on the Manifold of Symmetric Positive-Definite Matrices with Applications to Neuroimaging

Chun-Hao Yang[1] and Baba C. Vemuri[2](✉)

[1] Department of Statistics, University of Florida, Gainesville, FL, USA
[2] Department of CISE, University of Florida, Gainesville, FL, USA
vemuri@ufl.edu

Abstract. The James-Stein shrinkage estimator was proposed in the field of Statistics as an estimator of the mean for samples drawn from a Gaussian distribution and shown to dominate the maximum likelihood estimator (MLE) in terms of the risk. This seminal work lead to a flurry of activity in the field of shrinkage estimation. However, there has been very little work on shrinkage estimation for data samples that reside on manifolds. In this paper, we present a novel shrinkage estimator of the Fréchet Mean (FM) of manifold-valued data for the manifold, P_n, of symmetric positive definite matrices of size 'n'. We choose to endow P_n with the well known Log-Euclidean metric for its simplicity and ease of computation. With this choice of the metric, we show that the shrinkage estimator can be derived in an analytic form. Further, we prove that the shrinkage estimate of FM dominates the MLE of the FM in terms of the risk. We present several synthetic data examples with noise along with performance comparisons to estimated FM using other non-shrinkage estimators. As an application of shrinkage FM-estimation to real data, we compute the average motor sensory area (M1) tract from diffusion MR brain scans of controls and patients with Parkinson Disease (PD). We first show the dominance of the shrinkage FM estimator over the MLE of FM in this setting and then perform group testing to show differences between PD and controls based on the M1 tracts.

1 Introduction

In medical imaging, data taking the form of symmetric positive-definite (SPD) matrices are quite commonly encountered, for example, diffusion tensors, Cauchy deformation tensors, conductance tensors, etc. In such cases, data processing methods must perform geometry-aware computations, i.e., employ methods that take into account the nonlinear geometry of the data space. In medical imaging and many other domains, it is quite common to compute summary statistics from the data to characterize population groups. The most common and simple summary statistic is the mean. When the data space is nonlinear such as in the case of non-flat Riemannian manifolds, sample Fréchet mean (FM) is

A. C. S. Chung et al. (Eds.): IPMI 2019, LNCS 11492, pp. 566–578, 2019.
https://doi.org/10.1007/978-3-030-20351-1_44

the statistic we seek to compute. Sample FM is defined as the minimizer of the sum of squared geodesic distances from the data samples to the unknown center. This minimization is solved traditionally using the Riemannian gradient descent. Recently however, provably convergent and efficient recursive algorithms have been presented for computing the sample FM on a variety of Riemannian manifolds [4,10,17,22]. In the Euclidean space $\mathbb{R}^n$ with the usual metric, the sample FM is simply the sample mean of the observations and the James-Stein estimator [11], or a shrinkage estimator, is an estimator that is well known to be uniformly better (in terms of risk) than the sample mean when the observations are assumed to be normally distributed. Hence, the goal of this paper is to develop a novel shrinkage estimator for data residing on the space of SPD matrices.

The James-Stein estimator originated from the following problem. Given $X_i \overset{ind}{\sim} \mathcal{N}\left(\mu_i, \sigma^2\right)$, $i = 1, \ldots, p$ where $p > 2$ and σ^2 is known, what is a good estimator for μ_i under a quadratic loss? An intuitive answer would be the MLE X_i. However, Stein [20] proved that the MLE is inadmissible, and provided a class of estimators that dominate the MLE. Later, James and Stein [11] further sharpened the result and proposed the following estimator,

$$\left(1 - \frac{(p-2)\sigma^2}{\|X\|^2}\right) X_i \tag{1}$$

for μ_i, where $X = [X_1, \ldots, X_p]^T$. This estimator is referred to as the James-Stein estimator or the shrinkage estimator.

Ever since then, researchers have been trying to generalize this shrinkage phenomenon and apply it to different problems. For example, authors of [14] and [5] report a shrinkage estimator for a covariance matrix and authors of [3] and [24] generalized the shrinkage estimator to other family of distributions. From an applications perspective, authors of [15] developed a James-Stein version of Kalman filter which yielded robust parameter estimates in the presence of outliers in the data. In [16], authors proposed a shrinkage estimator to estimate the mean function in the Reproducing Kernel Hilbert space (RKHS). Shrinkage estimators for multi-task averaging problems was addressed recently in [7]. In [8], authors presented an interesting application of James-Stein estimation to the problem of geodesic regression in the space of diffeomorphisms to fit a generative model to images acquired over time. They showed that the shrinkage estimator of the momentum parameter estimated from cross-sectional scans and used to regularize the individual geodesic model improves prediction of the individual generative model. In all of the works cited above, the domain of the data has invariably been a vector space and as mentioned earlier, many applications naturally encounter data residing in non-Euclidean spaces. Hence, generalizing the shrinkage estimator to non-Euclidean spaces is a worthwhile pursuit. In this work, we focus on one such generalization of shrinkage estimation to the Riemannian manifold of SPD matrices.

In this paper, we derive a shrinkage estimator on the space of SPD matrices using a Bayesian framework for developing shrinkage estimators presented in

Xie et al. [25] and show that the proposed estimator is asymptotically optimal. We design synthetic experiments to demonstrate that the proposed estimator is better (in terms of risk) than the widely used Riemannian gradient descent based estimator and the recently developed inductive/recursive FM estimator in [10]. Further, we also apply the shrinkage estimator to find group differences between patients with Parkinson Disease and Controls.

Rest of this paper is organized as follows. In Sect. 2, we will present relevant background material about the space of SPD matrices and shrinkage estimation. The main result is presented in Sect. 3. Synthetic and real data experiments depicting the dominance of our shrinkage estimator of FM over MLE of FM are presented in Sect. 4. Finally, we conclude in Sect. 5.

2 Preliminaries

This section contains a review of some background differential geometry and statistics material that will be needed in the rest of the paper.

2.1 Geometry of P_n

We now present basic Riemannian geometry of symmetric positive definite (SPD) matrices denoted by P_n and refer the reader to [9,23] for details. The manifold P_n of $n \times n$ SPD matrices is defined as, $P_n = \{X = (x_{ij})_{1 \leq i,j \leq n} | X = X^T, \forall v \in \mathbb{R}^n, v \neq 0, v^T X v > 0\}$. The most commonly encountered example of SPD matrices is the covariance matrix (with non-zero eigenvalues), which is widely used in medical imaging, statistics, finance, computer vision and other fields. On P_n, the most widely used Riemannian metric is given by

$$\langle U, V \rangle_X = tr(X^{-1/2} U X^{-1} V X^{-1/2}) \tag{2}$$

for $X \in P_n$, $U, V \in T_X P_n$. The most important property of this metric the GL-invariance, i.e. for $g \in GL(n)$, $\langle U, V \rangle_X = \langle gUg^T, gVg^T \rangle_{gXg^T}$. Hereafter we refer this metric as GL-invariant metric to avoid confusion. With the GL-invariant metric, the induced geodesic distance between $X, Y \in P_n$ is given by (see [23])

$$d_{GL}(X, Y) = \sqrt{tr((\log(X^{-1}Y))^2)} \tag{3}$$

where log is the matrix logarithm. Since this distance is induced from the GL-invariant metric in Eq. (2), it is naturally GL-invariant i.e. $d_{GL}(X, Y) = d_{GL}(gXg^T, gYg^T)$.

More than a decade ago, Arsigny et al. [1] proposed the Log-Euclidean metric on the manifold P_n. This metric makes P_n a flat Riemannian manifold. The geodesic distance $d_{LE} : P_n \times P_n \to \mathbb{R}$ induced by the Log-Euclidean metric has a particularly simple form,

$$d_{LE}(X, Y) = \|\log X - \log Y\|_F .$$

Since for $X \in P_n$, $\log X \in \mathsf{Sym}(n) = \{X \in GL(n) | X = X^T\}$ and $\mathsf{Sym}(n)$ is isomorphic to $\mathbb{R}^{\frac{n(n+1)}{2}}$, it is convenient to use the map $vecd : \mathsf{Sym}(n) \to \mathbb{R}^{\frac{n(n+1)}{2}}$ defined in [19]. For $Y \in \mathsf{Sym}(n)$,

$$vecd(Y) = \left[y_{11}, ..., y_{nn}, \sqrt{2}(y_{ij})_{i<j}\right]^T.$$

For example,

$$Y = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{12} & y_{22} & y_{23} \\ y_{13} & y_{23} & y_{33} \end{bmatrix}, vecd(Y) = \begin{bmatrix} y_{11} & y_{22} & y_{33} & \sqrt{2}y_{12} & \sqrt{2}y_{13} & \sqrt{2}y_{23} \end{bmatrix}^T.$$

To simplify the notation, for $X \in P_n$, we denote $\widetilde{X} = vecd(\log X) \in \mathbb{R}^{\frac{n(n+1)}{2}}$. From the definition of $vecd$, it is easy to see that $d_{LE}(X, Y) = \|\widetilde{X} - \widetilde{Y}\|$.

Given $X_1, \ldots, X_N \in P_n$, we denote the sample FM with respect to the two geodesic distances given above by,

$$\bar{X}_N^{GL} = \underset{M \in P_n}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} d_{GL}^2(X_i, M) \; and \tag{4}$$

$$\bar{X}_N^{LE} = \underset{M \in P_n}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} d_{LE}^2(X_i, M) = \exp\left(\frac{1}{N} \sum_{i=1}^{N} \log X_i\right). \tag{5}$$

2.2 The Log-Normal Distribution on P_n

To model observations residing directly on P_n, Schwartzman [18] proposed the Log-Normal distribution which can be viewed as a generalization of the Log-Normal distribution on $\mathbb{R}^+$ to P_n.

Definition 1. *Let X be a P_n-valued random variable. We say X follows a Log-Normal distribution with mean $M \in P_n$ and covariance matrix $\Sigma \in P_{n(n+1)/2}$, or $X \sim LN(M, \Sigma)$ if*

$$\widetilde{X} \sim N(\widetilde{M}, \Sigma)$$

Important properties for this distribution are studied in [19]. The following proposition from [19] will be useful subsequently in this work. The proof of this proposition is straightforward and hence omitted.

Proposition 1. *Let $X_1, \ldots, X_N \overset{i.i.d}{\sim} LN(M, \Sigma)$. Then MLEs of M and Σ are $\hat{M}^{MLE} = \bar{X}_N^{LE}$ and $\hat{\Sigma}^{MLE} = \frac{1}{N}\sum_i \left(\widetilde{X}_i - \widetilde{\hat{M}^{MLE}}\right)\left(\widetilde{X}_i - \widetilde{\hat{M}^{MLE}}\right)^T$. The MLE of M is the sample FM under the Log-Euclidean metric.*

2.3 Bayesian Formulation of the Shrinkage Estimation in $\boldsymbol{R}^n$

The shrinkage estimator arose from a simultaneous estimation problem namely: estimate μ_i given $X_i \overset{ind.}{\sim} N(\mu_i, \sigma^2)$, where $i = 1, \ldots, p$, $p > 2$, σ^2 is known. The seminal work of James and Stein [11] showed that the information contained in X_j, $j \neq i$ can help to improve the estimation of μ_i. Later on, Efron and Morris [6] formulated the same problem using a Bayesian model and gave an empirical Bayes interpretation to the shrinkage estimator. The corresponding Bayesian hierarchical model is given below:

$$\begin{aligned} X_i|\theta_i &\overset{ind}{\sim} N(\theta_i, A), i = 1, \ldots, p \\ \theta_i &\overset{i.i.d}{\sim} N(\mu, \lambda) \end{aligned}$$

where, A is known and μ and λ are unknown. The maximum a posteriori (MAP) estimate for θ_i is given by,

$$\hat{\theta}_i^{\lambda,\mu} = \frac{\lambda}{\lambda + A} X_i + \frac{A}{\lambda + A}\mu. \tag{6}$$

The unknown parameters λ and μ can be estimated by empirical Bayes MLE (EBMLE) or an empirical Bayes method of moments (EBMOM). For the special case of $\mu = 0$, the EBMLE and EBMOM produce the same estimator which is the James-Stein estimator (1). A natural question would then arise: is there an optimal shrinkage estimator, i.e. how to estimate λ and μ such that they are optimal within such a class of estimators. The optimality here is defined in terms of the risk function, or the expected value of the loss function $R(\hat{\theta}, \theta) = E_{\boldsymbol{\theta}} L(\hat{\theta}(X), \theta)$, where $\hat{\theta}(X)$ is an estimator of θ based on the observation X. An estimator $\hat{\theta}$ of θ is said to be optimal if $R(\hat{\theta}, \theta) \leq R(\hat{\theta}^{\star}, \theta)$ for all θ. Hence, the optimal choice of λ and μ are given by,

$$\hat{\lambda}^{\text{opt}}, \hat{\mu}^{\text{opt}} = \arg\min_{\lambda,\mu} R(\hat{\boldsymbol{\theta}}^{\lambda,\mu}, \boldsymbol{\theta}).$$

where, $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_p]^T$ and $\hat{\boldsymbol{\theta}}^{\lambda,\mu} = [\hat{\theta}_1^{\lambda,\mu}, \ldots, \hat{\theta}_p^{\lambda,\mu}]^T$. However, since $R(\hat{\boldsymbol{\theta}}^{\lambda,\mu}, \boldsymbol{\theta})$ involves $\boldsymbol{\theta}$, this problem is ill-posed. Motivated by Stein's unbiased risk estimate (SURE) [21], we minimize the unbiased risk estimate $\text{SURE}(\lambda, \mu)$ instead of the risk where,

$$E_{\boldsymbol{\theta}} \left[\text{SURE}(\lambda, \mu)\right] = R(\hat{\boldsymbol{\theta}}^{\lambda,\mu}, \boldsymbol{\theta}).$$

Hence,

$$\hat{\lambda}^{\text{SURE}}, \hat{\mu}^{\text{SURE}} = \arg\min_{\lambda,\mu} \text{SURE}(\lambda, \mu)$$

This approach has been used to derive estimators for different models. For example, Xie et al. [25] derived the (asymptotically) optimal shrinkage estimator for heteroscedastic hierarchical model and their result is further generalized by [12] and [13].

3 Theory

We are now ready to present the main theoretical results of this paper involving a Bayesian formulation of the shrinkage estimator of M, the FM of Log-Normal distribution on P_n and a theorem on the dominance of our shrinkage estimator over the MLE of M on P_n endowed with the Log-Euclidean metric. The choice of Log-Euclidean metric here over other metrics is dictated by (i) computational efficiency of this metric over other choices and (ii) the existence of a closed form expression for the shrinkage estimator (to be derived here).

We model the data in this work as follows:

$$\begin{aligned} X_{ij}|M_i &\overset{ind}{\sim} \mathrm{LN}(M_i, A_i I), j = 1, \ldots, N \\ M_i &\overset{i.i.d}{\sim} \mathrm{LN}(\boldsymbol{\mu}, \lambda I), i = 1, \ldots, p \end{aligned}$$

where A_i's are known and $\boldsymbol{\mu}$ and λ are unknown. Our goal in this paper is to develop a shrinkage estimator for M_i which is better than the MLE, $\bar{X}_i^{LE} = \exp(N^{-1}\sum_j \log X_{ij})$, in terms of risk. Given the above model, the MAP estimate for M_i is given by,

$$\hat{M}_i^{\lambda,\boldsymbol{\mu}} = \exp\left(\frac{\lambda}{\lambda + A_i}\log \bar{X}_i^{\mathrm{LE}} + \frac{A_i}{\lambda + A_i}\log \boldsymbol{\mu}\right). \tag{7}$$

Let $\boldsymbol{M} = [M_1, \ldots, M_p]$ and $\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}} = [\hat{M}_1^{\lambda,\boldsymbol{\mu}}, \ldots, \hat{M}_p^{\lambda,\boldsymbol{\mu}}]$. Using the loss function, $l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}}, \boldsymbol{M}) = \frac{1}{p}\sum_i d_{\mathrm{LE}}^2(\hat{M}_i^{\lambda,\boldsymbol{\mu}}, M_i)$, the risk function becomes,

$$\begin{aligned} R(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}}, \boldsymbol{M}) &= E\left[\frac{1}{p}\sum_{i=1}^{p} d_{\mathrm{LE}}^2(\hat{M}_i^{\lambda,\boldsymbol{\mu}}, M_i)\right] \\ &= \frac{1}{p}\sum_{i=1}^{p}\frac{A_i}{(\lambda + A_i)^2}\left(A_i\|\log\boldsymbol{\mu} - \log M_i\|^2 + \frac{q\lambda^2}{N}\right) \end{aligned}$$

where $q = n(n+1)/2$. Since λ and $\boldsymbol{\mu}$ are unknown, our goal is to find the optimal λ and $\boldsymbol{\mu}$ in the sense that the risk is the smallest for all $\boldsymbol{M}$. Using the formalism given in Sect. 2.3 for approximating the risk function by SURE, we have,

$$\mathrm{SURE}(\lambda, \boldsymbol{\mu}) = \frac{1}{p}\sum_{i=1}^{p}\frac{A_i}{(\lambda + A_i)^2}\left(A_i\|\log \bar{X}_i^{\mathrm{LE}} - \log\boldsymbol{\mu}\|^2 + \frac{q(\lambda^2 - A_i^2)}{N}\right).$$

Hence, the choices of λ and $\boldsymbol{\mu}$ would be

$$\begin{aligned} \hat{\lambda}^{\mathrm{SURE}}, \hat{\boldsymbol{\mu}}^{\mathrm{SURE}} &= \arg\min_{\lambda,\boldsymbol{\mu}} \mathrm{SURE}(\lambda, \boldsymbol{\mu}) \\ &= \arg\min_{\lambda,\boldsymbol{\mu}} \frac{1}{p}\sum_{i=1}^{p}\frac{A_i}{(\lambda + A_i)^2}\left(A_i\|\log \bar{X}_i^{\mathrm{LE}} - \log\boldsymbol{\mu}\|^2 + \frac{q(\lambda^2 - A_i^2)}{N}\right). \end{aligned}$$

The proposed shrinkage estimator, SURE-FM, for M_i is

$$\hat{M}_i^{\text{SURE}} = \exp\left(\frac{\hat{\lambda}^{\text{SURE}}}{\hat{\lambda}^{\text{SURE}} + A_i} \log \bar{X}_i^{\text{LE}} + \frac{A_i}{\hat{\lambda}^{\text{SURE}} + A_i} \log \hat{\boldsymbol{\mu}}^{\text{SURE}}\right). \tag{8}$$

Since $\hat{\lambda}^{\text{SURE}}$, $\hat{\boldsymbol{\mu}}^{\text{SURE}}$ are the minimizers of $\text{SURE}(\lambda, \boldsymbol{\mu})$ instead of $R(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}}, \boldsymbol{M}) = E\left[l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}}, \boldsymbol{M})\right]$, we show in Theorem 1 that $\text{SURE}(\lambda, \boldsymbol{\mu})$ is a good approximation of $l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}}, \boldsymbol{M})$.

Theorem 1. *Assume that,*

(A) $\limsup_{p\to\infty} \frac{1}{p}\sum_i A_i^2 < \infty$
(B) $\limsup_{p\to\infty} \frac{1}{p}\sum_i A_i \|\log M_i\|^2 < \infty$
(C) $\limsup_{p\to\infty} \frac{1}{p}\sum_i \|\log M_i\|^{2+\delta} < \infty$ *for some* $\delta > 0$.

Then,

$$\sup_{\lambda>0, \|\log \boldsymbol{\mu}\| < \max_i \|\log \bar{X}_i^{LE}\|} |SURE(\lambda, \boldsymbol{\mu}) - l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}}, \boldsymbol{M})| \to 0$$

in probability as $p \to \infty$.

Proof. Let $\widetilde{X}_i^{\text{LE}} = vecd(\log \bar{X}_i^{\text{LE}})$, $\widetilde{\boldsymbol{\mu}} = vecd(\log \boldsymbol{\mu})$, $\widetilde{M}_i^{\lambda,\boldsymbol{\mu}} = vecd(\log \hat{M}_i^{\lambda,\boldsymbol{\mu}})$, $\widetilde{M}_i = vecd(\log M_i)$. Then,

$$\text{SURE}(\lambda, \boldsymbol{\mu}) = \sum_{j=1}^{q} \frac{1}{p}\sum_{i=1}^{p} \frac{A_i}{(\lambda + A_i)^2}\left(A_i\left((\widetilde{X}_i^{\text{LE}})_j - (\widetilde{\boldsymbol{\mu}})_j\right)^2 + \frac{\lambda^2 - A_i^2}{N}\right) = \sum_{j=1}^{q} \text{SURE}_j(\lambda, (\widetilde{\boldsymbol{\mu}})_j)$$

and

$$l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}}, \boldsymbol{M}) = \sum_{j=1}^{q} \frac{1}{p}\sum_{i=1}^{p}\left((\widetilde{M}_i^{\lambda,\boldsymbol{\mu}})_j - (\widetilde{M}_i)_j\right)^2 = \sum_{j=1}^{q} l_j.$$

Hence by Theorem 5.1 in [25] we have,

$$\begin{aligned}&\sup_{\lambda>0, \|\log \boldsymbol{\mu}\| < \max_i \|\log \bar{X}_i^{\text{LE}}\|} |\text{SURE}(\lambda, \boldsymbol{\mu}) - l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}}, \boldsymbol{M})| \\ &\le \sum_{j=1}^{q} \sup_{\lambda>0, \|\log \boldsymbol{\mu}\| < \max_i \|\log \bar{X}_i^{\text{LE}}\|} |\text{SURE}_j(\lambda, (\widetilde{\boldsymbol{\mu}})_j) - l_j| \to 0\end{aligned}$$

in probability as $p \to \infty$. □

In next theorem, we will show that our proposed shrinkage estimator is asymptotically optimal in the sense that its risk is asymptotically smaller than any other estimator of the form (7).

Theorem 2. *Assume that (A), (B), (C) in Theorem 1 hold. Then,*

$$\lim_{p\to\infty}[R(\hat{\boldsymbol{M}}^{SURE},\boldsymbol{M})-R(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}},\boldsymbol{M})]\le 0$$

Proof. Since

$$\begin{aligned}l(\hat{\boldsymbol{M}}^{\text{SURE}},\boldsymbol{M})-l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}},\boldsymbol{M}) &= l(\hat{\boldsymbol{M}}^{\text{SURE}},\boldsymbol{M})-\text{SURE}(\hat{\lambda}^{\text{SURE}},\hat{\boldsymbol{\mu}}^{\text{SURE}})\\ &+\ \text{SURE}(\hat{\lambda}^{\text{SURE}},\hat{\boldsymbol{\mu}}^{\text{SURE}})-\text{SURE}(\lambda,\boldsymbol{\mu})\\ &-\ l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}},\boldsymbol{M})+\text{SURE}(\lambda,\boldsymbol{\mu})\\ &\le\ 2\sup|\text{SURE}(\lambda,\boldsymbol{\mu})-l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}},\boldsymbol{M})|,\end{aligned}$$

from Theorem 1, we have

$$\lim_{p\to\infty}\left[l(\hat{\boldsymbol{M}}^{\text{SURE}},\boldsymbol{M})-l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}},\boldsymbol{M})\right]\le 0.$$

Hence,

$$\lim_{p\to\infty}\left[R(\hat{\boldsymbol{M}}^{\text{SURE}},\boldsymbol{M})-R(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}},\boldsymbol{M})\right]=\lim_{p\to\infty}E\left[l(\hat{\boldsymbol{M}}^{\text{SURE}},\boldsymbol{M})-l(\hat{\boldsymbol{M}}^{\lambda,\boldsymbol{\mu}},\boldsymbol{M})\right]\le 0.$$

□

4 Experiments

In this section, we present both synthetic and real data experiments to show that the SURE-FM is better than the MLE of the FM on P_n in terms of risk.

4.1 Synthetic Data Experiments

In this subsection, we will demonstrate the dominance of the SURE-FM over MLE of the FM of Log-Normal distribution on P_n using synthetically generated data. We compare the performance of the three different estimators namely, (i) SURE-FM, (ii) MLE of FM, denoted FM.LE and (iii) sample FM using the GL-invariant metric (using the recursive algorithm in [10]), denoted FM.GL. We use the following loss function in our comparisons of accuracy, $l(\hat{M},M)=d^2_{\text{LE}}(\hat{M},M)$. The lower the loss, the better the estimator. The procedure is shown in Algorithm 1 and in all of our experiments we set $m=1000$.

Algorithm 1. Procedure for synthetic data experiment on P_3.

```
Input: sample size $N$, variance $\lambda$, dimension $p$
Output: $R^{\text{LE}}$, $R^{\text{GL}}$, $R^{\text{SURE}}$
for $k = 1$ to $m$ do
    Generate $M_1, ..., M_p \overset{i.i.d}{\sim} \text{LN}(I, \lambda I)$
    Generate $A_1, ..., A_p \sim \text{Uniform}(1, 5)$
    Generate $X_{ij} \sim \text{LN}(M_i, A_i I)$, $j = 1, ..., n$, $i = 1, ..., p$
    Compute $\bar{X}_i^{GL}$, $\bar{X}_i^{LE}$, and $\hat{M}_i^{\text{SURE}}$ in (4), (5), and (8)
    Compute the loss $l_k^{\text{LE}} = l(\bar{X}^{\text{LE}}, \boldsymbol{M})$, $l_k^{\text{GL}} = l(\bar{X}^{\text{GL}}, \boldsymbol{M})$, and
    $l_k^{\text{SURE}} = l(\hat{\boldsymbol{M}}^{\text{SURE}}, \boldsymbol{M})$.
Compute $R^{\text{LE}} = \frac{1}{m}\sum_k l_k^{\text{LE}}$, $R^{\text{GL}} = \frac{1}{m}\sum_k l_k^{\text{GL}}$, and $R^{\text{SURE}} = \frac{1}{m}\sum_k l_k^{\text{SURE}}$.
```

In our experiments, we chose $\lambda = 1, 2, 4$, $n = 3, 5, 10, 20$, and $p = 50, 100, 150, 200$ to see how the performance changes under varying parameter values. The results are shown in Fig. 1. The percentages of improvement range from 20% to 40% under varying conditions. It is evident that the SURE-FM yields smaller average loss compared to the other two estimates of FM in most of the cases. In Fig. 2, we show the computational cost for different estimators. As discussed in [1], the Log-Euclidean metric based sample FM computation is much more efficient than the GL-invariant based sample FM computation. The SURE-FM is slightly slower than the FM.LE computation because of an extra optimization step that is involved.

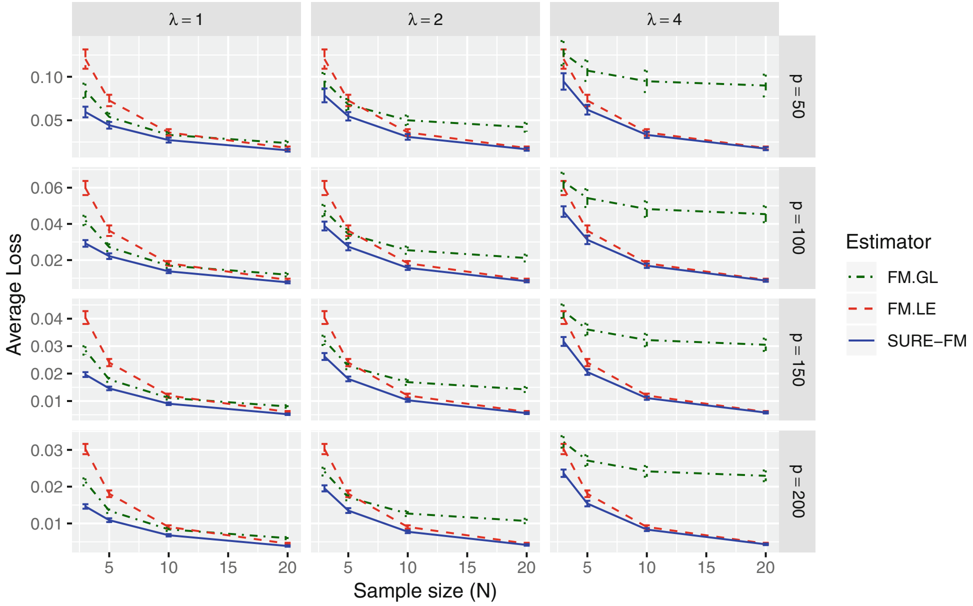

Fig. 1. The average loss for the three different estimators. Results for λ variation are shown across the columns and varying dimension p are shown across the rows.

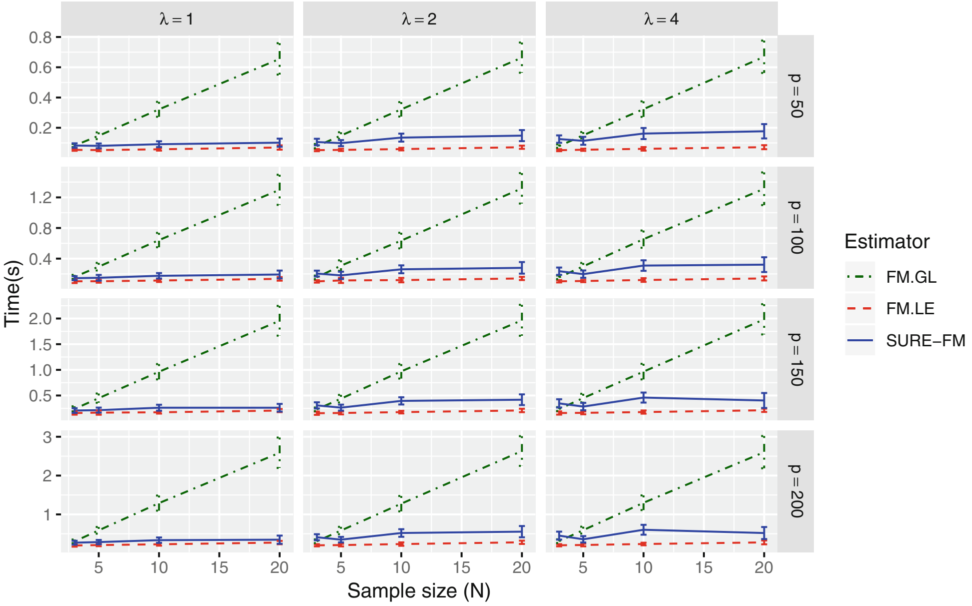

Fig. 2. The average time (on a log scale) taken for computing the three different estimators. Results for λ variation are shown across the columns and varying dimension p are shown across the rows.

4.2 Real Data Experiments

For the real data experiments, we test the performance of SURE-FM on the diffusion MRI datasets. The data consists of 50 patients with Parkinsons disease (PD) and 44 control cases (CON). The parameters of the diffusion image acquisition sequence were as follows: repetition time = 7748 ms, echo time = 86 ms, flip angle = 90, # of diffusion gradients: 64, field of view = 224 224 mm, in-plane resolution = 2 mm isotropic, slice-thickness = 2 mm, SENSE factor = 2.

We extract the motor sensory area fiber tracts (M1 fiber tracts) from each member of the two groups (PD and CON) using the FSL software [2] and each tract here spans across 33 voxels for the left hemisphere tract and 34 voxels for the right hemisphere tract respectively. We then fit diffusion tensors to each voxel along each of the tracts to obtain 33 (34) (3×3) SPD matrices. We then compute the FM tract for each group (CON and PD). The FM tract here also has 33 (34) diffusion tensors along the tract. We will use these FMs computed from the full population of each group as the 'ground truth'. Then, we randomly draw a subsample of size $N = 3, 5, 10, 20$ from each group (PD and CON) and compute the FM.LE, FM.GL, and SURE-FM of each group for the aforementioned subsample. We compare the performance of different estimators by the distance between the estimator and the 'ground truth' FMs. We repeat the experiment for $m = 1000$ random draws of subsamples and report the average distances. The results are shown in Table 1.

Table 1. The average distances from the subsample FMs and subsample SURE-FM to the population FM.

N	3	5	10	20
FM.LE	0.0827	0.0519	0.0231	0.0097
FM.GL	0.0814	0.0509	0.0224	0.0094
SURE-FM	**0.0738**	**0.0466**	**0.0211**	**0.0092**

The result shows that the SURE-FM dominates the MLE estimates of FM. As the sample size increases, the improvement is less significant which is consistent with the observations on synthetic data experiments in Sect. 4.1.

Finally, we apply the SURE-FM to find group differences between PD and CON data (described above) based on the M1 fiber tracts on both hemispheres of the brain. We use permutation testing to assess the group differences. The test statistic here is the difference of the SURE-FM of the two groups denoted by $d^{\text{SURE-FM}}$. We repeat the permutation step 10,000 times and recorded the differences of SURE-FM $d_i^{\text{SURE-FM}}$, $i = 1, \ldots, 10,000$. The p-value of 0.042 is obtained as a fraction of times that $d^{\text{SURE-FM}} < d_i^{\text{SURE-FM}}$. This low p-value is indicative of the significant difference found between the two groups using SURE-FM.

5 Discussion and Conclusions

In this paper, we presented a Bayesian formulation to generalize shrinkage estimation from $\mathbb{R}^n$ to the manifold of SPD matrices and proved that it dominates the MLE of the FM in terms of risk The shrinkage factor and the shrinkage target are obtained by minimizing the Stein's unbiased risk estimate (SURE). Our theoretical results were derived using the Log-Euclidean metric, which is easy to compute and easy to manipulate formulae in our quest for the shrinkage estimator on P_n. We showed experimentally on synthetic and real data that SURE-FM is better than the sample FM estimates computed using the Log-Euclidean and the GL-invariant metrics respectively. The experiments depicted the dominance of the SURE-FM over MLE estimates as expected in the small sample size scenarios. This scenario is very pertinent to the medical imaging domain where one is faced with small sample population size but very high dimensional feature spaces. Thus, we envision that the research reported here can prove to be quite useful for statistical inference in such settings.

As is well known, the Log-Euclidean metric is not affine invariant and in some applications, such a property might be useful. However, from our preliminary attempts, we found it to be very challenging and almost intractable to derive a closed form solution for the estimator. Our future efforts will therefore focus on using symbolic manipulation tools to explore the possibility of tackling this problem. In parallel, we are also exploring formulations of shrinkage estimators for other manifolds commonly encountered in medical imaging applications.

Acknowledgements. This research was in part funded by the NSF grants IIS-1525431 and IIS-1724174 to BCV.

References

1. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. **29**(1), 328–347 (2007)
2. Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage **34**(1), 144–155 (2007)
3. Brandwein, A.C., Strawderman, W.E.: Stein estimation for spherically symmetric distributions: recent developments. Stat. Sci. **27**(1), 11–23 (2012)
4. Chakraborty, R., Vemuri, B.C.: Recursive fréchet mean computation on the grassmannian and its applications to computer vision. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4229–4237 (2015)
5. Daniels, M.J., Kass, R.E.: Shrinkage estimators for covariance matrices. Biometrics **57**(4), 1173–1184 (2001)
6. Efron, B., Morris, C.: Stein's estimation rule and its competitorsan empirical Bayes approach. J. Am. Stat. Assoc. **68**(341), 117–130 (1973)
7. Feldman, S., Gupta, M.R., Frigyik, B.A.: Revisiting stein's paradox: multi-task averaging. J. Mach. Learn. Res. **15**(1), 3441–3482 (2014)
8. Fleishman, G.M., Fletcher, P.T., Gutman, B.A., Prasad, G., Wu, Y., Thompson, P.M.: Geodesic refinement using james-stein estimators. Math. Found. Comput. Anat. **60**, 60–70 (2015)
9. Helgason, S.: Differential Geometry, Lie Groups and Symmetric Spaces. American Mathematical Society, Providence (2001)
10. Ho, J., Cheng, G., Salehian, H., Vemuri, B.: Recursive Karcher expectation estimators and geometric law of large numbers. In: Artificial Intelligence and Statistics, pp. 325–332 (2013)
11. James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 361–379 (1961)
12. Jing, B.Y., Li, Z., Pan, G., Zhou, W.: On sure-type double shrinkage estimation. J. Am. Stat. Assoc. **111**(516), 1696–1704 (2016)
13. Kong, X., Liu, Z., Zhao, P., Zhou, W.: Sure estimates under dependence and heteroscedasticity. J. Multivar. Anal. **161**, 1–11 (2017)
14. Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. J. Multivar. Anal. **88**(2), 365–411 (2004)
15. Manton, J.H., Krishnamurthy, V., Poor, H.V.: James-stein state filtering algorithms. IEEE Trans. Sig. Process. **46**(9), 2431–2447 (1998)
16. Muandet, K., Sriperumbudur, B., Fukumizu, K., Gretton, A., Schölkopf, B.: Kernel mean shrinkage estimators. J. Mach. Learn. Res. **17**(1), 1656–1696 (2016)
17. Salehian, H., Chakraborty, R., Ofori, E., Vaillancourt, D., Vemuri, B.C.: An efficient recursive estimator of the fréchet mean on a hypersphere with applications to medical image analysis. Math. Found. Comput. Anat. **3**, 143–154 (2015)
18. Schwartzman, A.: Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data. Ph.D. thesis, Stanford University (2006)
19. Schwartzman, A.: Lognormal distributions and geometric averages of symmetric positive definite matrices. Int. Stat. Rev. **84**(3), 456–486 (2016)
20. Stein, C.: Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. 1, pp. 197–206. University of California Press, Berkeley (1956)

21. Stein, C.M.: Estimation of the mean of a multivariate normal distribution. Ann. Stat. **9**(6), 1135–1151 (1981)
22. Sturm, K.T.: Probability measures on metric spaces of nonpositive. In: Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces. Lecture Notes from a Quarter Program on Heat Kernels, Random Walks, and Analysis on Manifolds and Graphs, 16 April–13 July 2002, vol. 338, p. 357. Emile Borel Centre of the Henri Poincaré Institute, Paris (2003)
23. Terras, A.: Harmonic Analysis on Symmetric Spaces and Applications II. Springer Science & Business Media, New York (2012)
24. Xie, X., Kou, S.C., Brown, L.: Optimal shrinkage estimation of mean parameters in family of distributions with quadratic variance. Ann. Stat. **44**(2), 564 (2016)
25. Xie, X., Kou, S., Brown, L.D.: Sure estimates for a heteroscedastic hierarchical model. J. Am. Stat. Assoc. **107**(500), 1465–1479 (2012)