Statistica Sinica **31** (2021), 1829-1852
doi:https://doi.org/10.5705/ss.202019.0339

NEW WAVELET SURE THRESHOLDS OF ELLIPTICAL DISTRIBUTIONS UNDER THE BALANCE LOSS

Hamid Karamikabir and Mahmoud Afshari

Persian Gulf University

Abstract: In this paper, we introduce a new shrinkage soft-wavelet threshold estimator based on Stein's unbiased risk estimate (SURE) for elliptical and spherical distributions under balanced loss functions. we focus on particular thresholding rules to obtain a new threshold, and thus produce new estimators. In addition, we obtain SURE shrinkage based on nonnegative subset of the mean vector. Finally, we present a simulation to test the validity of the proposed estimator.

Key words and phrases: Balance loss function, elliptically distribution, restricted estimator, shrinkage estimator, soft wavelet estimator, spherically distribution.

1. Introduction

Mean vector (location) parameter estimation is an important problem in the context of point estimation. In a point estimation, we want an estimator that minimizes the risk function for all possible parameter values. However, this is not possible in practice, given the size of the class of estimators. To evaluate the performance of an estimator, we need to set a measure. Location vector parameter estimation is can be performed using a variety of loss functions, one of which is the balanced loss function. In this study, we use the balanced loss function in the following two ways.

Definition 1. Suppose that X is a random vector with mean vector parameter θ and scalar variational component Σ. The balanced error loss function, BEL(δ_0), is defined as

$$L_{\omega,\delta_0}(\theta,\delta) = \omega\left(\delta-\delta_0\right)^T \Sigma^{-1}\left(\delta-\delta_0\right) + (1-\omega)\left(\delta-\theta\right)^T \Sigma^{-1}\left(\delta-\theta\right), \qquad (1.1)$$

where $0 \le \omega < 1$, δ_0 is a target estimator, and Σ is symmetric nonsingular scale matrix.

As a special case, suppose that X is a random vector with mean vector parameter θ and scalar variational component σ^2 (i.e., $\Sigma = \sigma^2 I_p$ in Definition 1).

Corresponding author: Mahmoud Afshari, Department of Statistics, Persian Gulf University, Bushehr, Iran. E-mail: afshar@pgu.ac.ir.

Then, the balanced error loss function is as follows:

$$L_{\omega,\delta_0}(\theta,\delta) = \omega\frac{\|\delta-\delta_0\|^2}{\sigma^2} + (1-\omega)\frac{\|\delta-\theta\|^2}{\sigma^2}, \quad 0 \le \omega < 1. \tag{1.2}$$

The corresponding risk is the expectation with respect to the loss function. Then, the associated risk function with respect to (1.2) is $R(\theta,\delta) = E_\theta\left[L(\theta,\delta)\right]$. A special case of the balanced error loss function is the weighted quadratic loss when $\omega = 0$. The balanced loss function was introduced by Zellner (1994) to reflect two criteria: goodness of fit and precision of estimation. For further information on the use of this loss, refer to Jafari Jozani, Marchand and Parsian (2006), Cao and He (2017) and Karamikabir, Afshari and Arashi (2018).

Shrinking and truncating either the data directly or the coefficients in their Fourier series expansions is an popular technique in signal processing. For nonlocal bases, such as the trigonometric bases, shrinking the coefficients can affect the global shape of the reconstructed function and introduce unwanted artifacts. In the context of function estimation using wavelets, the shrinkage has an additional feature; it is related to smoothing (denoising), because the measure the of smoothness of a function depends on the magnitudes of its wavelet coefficients (Vidakovic (2009)).

Shrinkage estimation improved a raw estimator, in some sense, by combining it with other information. Although the shrinkage estimator is biased, it is well known that it has a minimum quadratic risk compared to natural estimators (mostly, the maximum likelihood estimator). The general form of the shrinkage estimator is $X + g(X)$. The shrinkage is usually performed by decreasing X using $g(X)$. Although in everyday life the notion of shrinkage may carry a negative connotation, it is not so in the domain of statistical estimation. Many good estimators are some sort of shrinkage estimators. For example, most Bayesian, minimax, and Gamma-minimax estimators are shrinkage estimators.

Donoho and Johnstone (1995) developed a technique for selecting a threshold by minimizing Stein's unbiased risk estimator (SURE) (Stein (1981)). This threshold is implemented in an adaptive denoising procedure, SureShrink. The adaptation in SureShrink is achieved by specifying thresholds in a level-wise manner.

One way to find an estimator for the parameter vector is to use the shrinkage wavelet method. Hard thresholding and soft-thresholding are two examples of shrinkages. The soft-thresholding wavelet shrinkage estimator of Antoniadis (2007) is given by the following:

$$\delta_\lambda^{soft}(X_i) = (X_i - sign(X_i)\lambda)I(|X_i| > \lambda) = \begin{cases} X_i + \lambda, & X_i < -\lambda, \\ 0, & |X_i| \leq \lambda, \\ X_i - \lambda, & X_i > \lambda, \end{cases}$$

where $I(\cdot)$ is an indicator function. Suppose that $X = (X_1, X_2, \dots, X_p)$. In this case, the $\delta_\lambda^{soft}(X)$ estimator can be written as

$$\delta_\lambda^{soft}(X) = X + g(X), \tag{1.3}$$

where $g(X) = (g(X_1), g(X_2), \dots, g(X_p))$, and $g(X_i)$ is defined as

$$g(X_i) = \begin{cases} \lambda, & X_i < -\lambda, \\ -X_i, & |X_i| \leq \lambda, \\ -\lambda, & X_i > \lambda. \end{cases} \tag{1.4}$$

Thus far, we have considered using different types of thresholds as shrinkage wavelets, such as the universal $\lambda^U = \sqrt{2 \log p}$ (p is the dim of the parameter vector), percentile (Mallat (1989)), cross-validation (CV) (Nason (1996)), false-discovery-rate (Benjamini and Hochberg (1995)), Lorentz (Lorentz (1905)), and block thresholds (Cai and Silverman (2001)), as well as the SURE (Donoho and Johnstone (1995)).

In this study, we generalize the result of Donoho and Johnstone (1995) for SURE thresholds with changes in the class of family distributions and the loss function. For this purpose, we derive new estimators for the location parameter of the multivariate normal, elliptical, or spherically symmetric distribution in classes of shrinkage estimators that are well behaved under a balanced error loss. These results generalize those of Donoho and Johnstone (1995), Fourdrinier and Ouassou (2000), Fourdrinier, Ouassou and Strawderman (2003), and Fourdrinier and Strawderman (2015) on the shrinkage wavelet approach. We derive estimates of the risk using Stein's lemma of the risk of those estimators. Furthermore, we derive expressions for the optimal λ under the SURE in the multivariate normal, spherically symmetric, and elliptical cases. In addition, we find the SURE shrinkage in the restricted parameter space. Here, we generalize previous works by Fourdrinier, Ouassou and Strawderman (2003) and Karamikabir, Afshari and Arashi (2018) on the shrinkage wavelet approach. In this respect, Chang and Strawderman (2017) studied a shrinkage estimation of p positive normal means under the sum of the squared error loss. More, recently Afshari, Lak and Gholizadeh (2017), Krebs (2018) and Karamikabir and Afshari (2019) applied wavelets for multivariate distributions. For more information refer to Van Eeden (2006) and Karamikabir, Afshari and Arashi (2018), among others.

Shrinkage wavelets have many applications in statistical estimation, some of which lie on the boundaries between disciplines. For instance, recent applications of wavelets to shape analysis are bridging statistical modeling, statistical theory of shapes, computational geometry, and image processing; see Vidakovic (2009).

Parameter estimation is important in statistical inference, especially when the goal is to estimate the parameter vector in a multivariate distribution. A well-known class of multivariate distributions is that of elliptical distributions.

Definition 2. The $p \times 1$ random vector $X = (X_1, \ldots, X_p)^T$ is said to have an elliptical distribution, denoted by $E_p(\theta, \Sigma, \psi)$, with $p \times 1$ vector location parameters θ, $p \times p$ scale matrix Σ, and the characteristic generator ψ, if its density function is of the form

$$C_m |\Sigma|^{-1/2} f\left((X-\theta)^T \Sigma^{-1} (X-\theta)\right),$$

where C_m is a normalizing constant. Its characteristic function satisfies

$$\phi(t) = E(e^{it^T X}) = e^{it^T \theta} \psi\left(t^T \Sigma t\right).$$

Examples of elliptical distributions include the multivariate normal distribution $N_p(\theta, \Sigma)$, multi-uniform distribution, multivariate Pearson type-II and type-VII distributions, multivariate Laplace distribution, generalized slash distribution, multivariate Cauchy distribution, multivariate Bessel distribution, multivariate exponential power distribution, and multivariate Kotz distribution.

A $p \times 1$ random vector X is said to have a spherically symmetric distribution (or simply a spherical distribution) if X and ΛX have the same distribution for all $p \times p$ orthogonal matrices Λ. The elliptical family are spherically symmetric distributions with diagonal scale parameter $\sigma^2 I_p$, and are represented as $SS(\theta, \sigma^2 I_p, \psi)$, where I_p is the identity matrix. Examples of spherical distributions include the multivariate normal distribution $N_p(0, \sigma^2 I_p)$, "ε-contaminated" normal distribution, multivariate t distribution, and scale mixture of multivariate normal distributions.

Donoho and Johnstone (1995) found the SURE threshold in a multivariate normal distribution $N_p(\theta, I_p)$. However, thus far, no one has investigated the SURE threshold for the spherical and elliptical distributions. There has also been no research on the restricted parameter space. In this study, we attempt to generalize this topics.

The reminder of the paper process as follows. In Section 2, we find a threshold based on the SURE under a balanced loss function in the class of elliptical and spherical distributions. In Section 3, we discuss the SURE shrinkage in the restricted parameter space, and in Section 4 the numerical performance of the proposed estimator using a simulation study. Section 5 concludes the paper.

2. A Threshold Based on the SURE

In this section, we first review the method of Donoho and Johnstone (1995) for finding the SURE threshold in a multivariate normal distribution $N_p(\theta, I_p)$. In the next step, we generalize the multivariate distribution and the loss function of the previous stage. In this way, instead of a multivariate normal distribution $N_p(\theta, I_p)$, we consider a multivariate normal distribution with a diagonal scale matrix $\sigma^2 I_p$, that is, $N_p(\theta, \sigma^2 I_p)$, a spherically symmetric distribution $SS(\theta, \sigma^2 I_p, \psi)$, and an elliptical distribution $E_p(\theta, \Sigma, \psi)$. In all three cases, instead of a general quadratic loss function we consider a balanced loss function.

Lemma 1. *(Stein (1981)) Suppose that $X \sim N_p(\theta, \sigma^2 I_p)$, with known σ^2. Then*

$$E_\theta[(X-\theta)^T g(X)] = \sigma^2 E[\nabla \cdot g(X)],$$

where $g : \mathbb{R}^p \to \mathbb{R}^p$ is a function for which the two expectations $E_\theta[(X-\theta)^T g(X)]$ and $E[\nabla \cdot g(X)]$ both exist, and $\nabla \cdot g(X)$ is the divergence operator with respect to the variable X:

$$\nabla \cdot g(X) = \sum_{i=1}^{p} \frac{\partial}{\partial X_i} g_i(X).$$

Donoho and Johnstone (1995) developed a technique for selecting a threshold by minimizing the SURE. This threshold is implemented in an adaptive denoising procedure, called SureShrink. The adaptation in SureShrink is achieved by using a level-wise specification. Specifically, suppose that $\theta = (\theta_1, \ldots, \theta_p)^T$ and $X_i \sim N(\theta_i, 1)$, for $i = 1, \ldots, p$. Let $\delta(X)$ be an estimator of θ. If the function $g = \{g_i\}_{i=1}^p$ in the shrinkage estimator representation $\delta(X) = X + g(X)$ is weakly differentiable (i.e., if there exists a function $q_\theta(x)$ such that $\partial/\partial\theta E_\theta[\delta(X)] = E[\delta(q(X))]$), then under the quadratic loss function $L(\theta, \delta) = \|\delta(X) - \theta\|^2$, by Lemma 1, we have the following expectation:

$$R(\theta, \delta(X)) = E_\theta\left(\|\delta(X)-\theta\|^2\right) = p + E_\theta\left(\|g(X)\|^2\right) + E_\theta\left(2\nabla \cdot g(X)\right). \tag{2.1}$$

We consider the soft-threshold shrinkage estimator $\delta_\lambda^{soft}(X) = X + g(X)$ in (1.3), where $g(X)$ is given in (1.4). We need the following useful equations for $g(X)$:

$$\nabla \cdot g(X) = -\sum_{i=1}^{p} I(|X_i| \leq \lambda), \quad \|g(X)\|^2 = \sum_{i=1}^{p} (|X_i| \wedge \lambda)^2. \tag{2.2}$$

where $(|X_i| \wedge \lambda) = min(|X_i|, \lambda)$. Because $\delta_\lambda^{soft}(X)$ is weakly differentiable in Stein's sense, and using Lemma 1 and Equation (2.2), we have from (2.1) that the quantity

$$SURE(X, \lambda) = p - 2\sum_{i=1}^{p} I(|X_i| \leq \lambda) + \sum_{i=1}^{p} (|X_i| \wedge \lambda)^2$$

is an unbiased estimate of risk; that is, $E_\theta(\|\delta_\lambda^{soft}(X) - \theta\|^2) = E_\theta\left[SURE\left(X, \lambda\right)\right]$. Consider using this estimator to select a threshold:

$$\lambda^{sure} = \underset{0 \leq \lambda \leq \lambda^U}{\operatorname{argmin}} \, SURE\left(X, \lambda\right).$$

Arguing heuristically, one would expect that, for a large dimension X, a sort of statistical regularity would set in, the law of large numbers would ensure that the SURE is close to the true risk, and λ^{SURE} would be almost the optimal threshold for the case at hand.

The λ^{sure} threshold and the soft-thresholding rule are the core of the level-dependent procedure of Donoho and Johnstone (1995), called SureShrink. If the wavelet representation at a particular level is not sparse, the SURE threshold is used.

2.1. Multivariate normal distribution with a diagonal scale matrix

Now, under the above condition, we consider the balanced loss function given in (1.2). In this case, similarly to Jafari Jozani, Marchand and Parsian (2006) and Karamikabir, Afshari and Arashi (2018), the target estimator can be part of the soft-threshold shrinkage estimator $\delta_\lambda^{soft}(X)$. The target estimator is as follows:

$$\delta_0(X) = X + (1 - \omega)g(X). \tag{2.3}$$

Hence the soft-threshold shrinkage estimator is $\delta_\lambda^{soft}(X) = \delta_0(X) + \omega g(X)$.

Theorem 1. *Suppose that $X \sim N_p(\theta, \sigma^2 I_p)$ with known σ^2. For the soft-threshold shrinkage estimator $\delta_\lambda^{soft}(X)$ and target estimator $\delta_0(X)$, the value of the threshold is given by*

$$\lambda^{sure} = \underset{0 \leq \lambda \leq \lambda^U}{\operatorname{argmin}} \; SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X), \delta_0(X)),$$

where $SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X), \delta_0(X))$ is given as follows:

$$p(1-\omega)+\frac{(\omega^3-\omega+1)}{\sigma^2}\sum_{i=1}^{p}(|X_i|\wedge\lambda)^2-\frac{2(1-\omega)}{\sigma^2}\sum_{i=1}^{p}I(|X_i|<\lambda).$$

Proof. By Lemma 1 and under BEL(δ_0) in (1.2), $R_{\omega,\delta_0}(\theta,\delta_\lambda^{soft}(X))$ is given as follows:

$$\begin{aligned}
&E\left[L_{\omega,\delta_0}(\theta,\delta_\lambda^{soft}(X))\right]\\
&=\frac{1}{\sigma^2}E\left(\omega\|\delta_\lambda^{soft}(X)-\delta_0(X)\|^2+(1-\omega)\|\delta_\lambda^{soft}(X)-\theta\|^2\right)\\
&=\frac{1}{\sigma^2}E\Big(\omega\|X+g(X)-X-(1-\omega)g(X)\|^2\\
&\quad+(1-\omega)\|X+g(X)-\theta\|^2\Big)\\
&=\frac{1}{\sigma^2}E\Big(\omega^3\|g(X)\|^2+(1-\omega)\|X-\theta\|^2+(1-\omega)\|g(X)\|^2\\
&\quad+2(1-\omega)(X-\theta)^Tg(X)\Big)\\
&=\frac{1}{\sigma^2}E\Big((\omega^3-\omega+1)\|g(X)\|^2+(1-\omega)\|X-\theta\|^2+2(1-\omega)\nabla\cdot g(X)\Big).
\end{aligned}$$

Then, $SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X),\delta_0(X))$ is equal to

$$p(1-\omega)+\frac{(\omega^3-\omega+1)}{\sigma^2}\sum_{i=1}^{p}(|X_i|\wedge\lambda)^2-\frac{2(1-\omega)}{\sigma^2}\sum_{i=1}^{p}I(|X_i|<\lambda),$$

and $SURE_{(X,\lambda)}$ is an unbiased estimate of risk, that is,

$$E\left(L_{\omega,\delta_0}(\theta,\delta_\lambda^{soft}(X))\right)=E\left[SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X),\delta_0(X))\right],$$

and $\lambda^{sure}=\mathrm{argmin}_{0\le\lambda\le\lambda^U}SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X),\delta_0(X))$.

Corollary 1. *For $\omega=0$ and $\sigma^2=1$, we have the result of Donoho and Johnstone (*1995*) for the SURE threshold.*

2.2. Spherically symmetric distribution

In this subsection, we consider unimodal spherically distributions and look for a SURE threshold under the balanced loss function. Assume (X,U) is a $p+k$ random vector that follows a spherically symmetric distribution around the $p+k$ vector $(\theta,0)$, where $\dim X=\dim\theta=p$, and $\dim U=\dim 0=k$, with spherically symmetric density

$$f_{X,U}(x,u) = \frac{1}{\sigma^{p+k}} f\left(\frac{\|x-\theta\|^2 + u^T u}{\sigma^2}\right), \tag{2.4}$$

where $\sigma \in \mathbb{R}_+$. Furthermore, suppose that the scalar variational component σ^2 is known. Under these conditions, we have the following lemma.

Lemma 2. *(Fourdrinier and Strawderman (1996)) For every weakly differentiable function $g : \mathbb{R}^p \to \mathbb{R}^p$, integer m, and $\theta \in \mathbb{R}^p$, we have*

$$E_\theta[(U^T U)^m g(X)^T (X-\theta)] = \frac{1}{k+2m} E_\theta[(U^T U)^{m+1} \nabla \cdot g(X)],$$

provided these expectations exist.

The shrinkage estimator introduced in Fourdrinier and Ouassou (2000) is $\delta(X,U) = X + U^T U g(X)$, where $g(\cdot)$ is some measurable function from $\mathbb{R}^p$ into $\mathbb{R}^p$, and the soft-threshold shrinkage estimator can be written as follows:

$$\delta_\lambda^{soft}(X,U) = X + U^T U g(X).$$

Let the target estimator be

$$\delta_0(X,U) = X + (1-\omega) U^T U g(X).$$

Hence we can write $\delta_\lambda^{soft}(X,U) = \delta_0(X,U) + \omega U^T U g(X)$.

Theorem 2. *Suppose that $(X,U) \sim SS_p((\theta,0), \sigma^2 I_p)$ with known σ^2. For the soft-threshold shrinkage estimator $\delta_\lambda^{soft}(X,U)$ and target estimator $\delta_0(X,U)$, the threshold λ^{sure} is given by*

$$\lambda^{sure} = \underset{0 \le \lambda \le \lambda^U}{\operatorname{argmin}}\ SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X,U), \delta_0(X,U)),$$

where $SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X,U), \delta_0(X,U))$ is given as follows:

$$\begin{aligned} &\frac{(\omega^3 - \omega + 1)\left(U^T U\right)^2}{\sigma^2} \sum_{i=1}^p (|X_i| \wedge \lambda)^2 \\ &+ p(1-\omega) - \frac{2(1-\omega)}{\sigma^2(k+2)} \left(U^T U\right)^2 \sum_{i=1}^p I(|X_i| \le \lambda). \end{aligned}$$

Proof. By substituting $m = 1$ into Lemma 2, we have the following risk:

$$\begin{aligned} R_{\omega,\delta_0}(\theta, \delta_\lambda^{soft}) &= E\left[L_{\omega,\delta_0}(\theta, \delta_\lambda^{soft}(X,U))\right] \\ &= \frac{1}{\sigma^2} E\Big(\omega \|\delta_\lambda^{soft}(X,U) - \delta_0(X,U)\|^2 \end{aligned}$$

$$
\begin{aligned}
&+(1-\omega)\|\delta_\lambda^{soft}(X,U)-\theta\|^2\Big)\\
&= \frac{1}{\sigma^2}E\Big(\omega\|X+U^TUg(X)-X-(1-\omega)U^TUg(X)\|^2\\
&\quad +(1-\omega)\|X+U^TUg(X)-\theta\|^2\Big)\\
&= \frac{1}{\sigma^2}E\Big[\omega^3\left(U^TU\right)^2\|g(X)\|^2\\
&\quad +(1-\omega)\Big(\|X-\theta\|^2+\left(U^TU\right)^2\|g(X)\|^2\\
&\quad +2(X-\theta)^TU^TUg(X)\Big)\Big]\\
&= \frac{1}{\sigma^2}E\Bigg[\left(\omega^3-\omega+1\right)\left(U^TU\right)^2\|g(X)\|^2\\
&\quad +(1-\omega)\left(\|X-\theta\|^2+\frac{2}{k+2}\left(U^TU\right)^2\nabla\cdot g(X)\right)\Bigg].
\end{aligned}
$$

Thus, $SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X,U),\delta_0(X,U))$ is given as follows:

$$
\begin{aligned}
&\frac{1}{\sigma^2}\frac{\left(\omega^3-\omega+1\right)\left(U^TU\right)^2}{\sigma^2}\sum_{i=1}^{p}(|X_i|\wedge\lambda)^2\\
&+p(1-\omega)-\frac{2(1-\omega)}{\sigma^2(k+2)}\left(U^TU\right)^2\sum_{i=1}^{p}I(|X_i|\le\lambda).
\end{aligned}
$$

Furthermore, the $SURE_{(X,\lambda)}$ is an unbiased estimate of risk, that is,

$$
E\left[L_{\omega,\delta_0}(\theta,\delta_\lambda^{soft}(X,U))\right]=E\left[SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X,U),\delta_0(X,U))\right],
$$

and $\lambda^{sure}=\mathrm{argmin}_{0\le\lambda\le\lambda^U}SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X,U),\delta_0(X,U))$.

2.3. Elliptical distribution

In this subsection, the statistical distribution is of the elliptical contoured distribution in Fourdrinier, Ouassou and Strawderman (2003).

Let $(X,V)=(X,V_1,\ldots,V_n)$ be an $n+1$ random vector in $\mathbb{R}^p$ with an elliptically contoured distribution of the form

$$
f(x,v)=\left|\Sigma^{-1}\right|f\left((x-\theta)^T\Sigma^{-1}(x-\theta)+\sum_{i=1}^{n}v_i^T\Sigma^{-1}v_i^T\right), \tag{2.5}
$$

where X and V_i are $p\times 1$ vectors, θ is a $p\times 1$ unknown location vector, and Σ is a $p\times p$ known matrix proportional to the covariance matrix. This density arises as a joint density in the canonical form of the general linear model. Here X is a projection on the space spanned by θ, and V_i is a projection onto the orthogonal complement of the space spanned by θ or $X \sim E(\theta, \Sigma, \psi)$ and $V = (V_1, \ldots, V_n) \sim E(0, \Sigma, \psi)$. The class in (2.5) contains models such as the multivariate normal, t, and Kotz-type distributions.

In the case of a normal distribution, (i.e., $X \sim N_p(\theta, \Sigma)$, $V = (V_1, \ldots, V_n)$, where $V_i \sim N_p(0, \Sigma)$ denotes an independent and $S = VV^T \sim W_p(n, \Sigma)$ (Wishart distribution) with $n \geq p$), the following lemma is a straightforward extension of Stein (1981) lemma and of Haff (1979) lemma.

Lemma 3. *(Fourdrinier and Strawderman (*2015*)) Assume that* $(X, V) = (X, V_1, \ldots, V_n)$ *is an* $n+1$ *random vector in* $\mathbb{R}^p$ *with a multivariate normal distribution, where* $X \sim N_p(\theta, \Sigma)$*, and* $V = (V_1, \ldots, V_n)$*,* $V_i \sim N_p(0, \Sigma)$*. In addition, let* $g(X, S)$ *be a* $p \times 1$ *vector, such that the function* $g(X, \cdot)$ *is weakly differentiable. Then we have*

$$E_\theta \left[(X - \theta)^T \Sigma^{-1} g(X, S) \right] = E_\theta \left[\nabla \cdot g(X, S) \right], \tag{2.6}$$

provided the expectations in (2.6) *exist. As defined before,* $\nabla \cdot g(X, S)$ *is the divergence operator with respect to the variable* X*.*

Similarly to Fourdrinier, Strawderman and Wells (2003), we define the expectations E_θ^* with respect to the distribution $C^{-1} F((X - \theta)^T \Sigma^{-1} (X - \theta) + \sum_{j=1}^n V_j^T \Sigma^{-1} V_j)$, where F and C are defined as follows:

$$F(t) = \frac{1}{2} \int_t^\infty f(s) ds,$$

$$C = \int_{\mathbb{R}^p \times \cdots \times \mathbb{R}^p} F \left((x - \theta)^T \Sigma^{-1} (x - \theta) + \sum_{j=1}^n v_j^T \Sigma^{-1} v_j \right) dx dv_1 \cdots dv_n.$$

Lemma 4. *(Fourdrinier, Strawderman and Wells (*2003*)) Let* (X, V) *be an* $n+1$ *random vector in* $\mathbb{R}^p$ *with an elliptically contoured distribution, and let* $S = VV^T$*. Assume that* $g(X, S)$ *is a function on* $\mathbb{R}^p$ *that is weakly differentiable in* X *and differentiable in* S*. Then,*

1. $E_\theta \left[g^T(X, S) \Sigma^{-1} (X - \theta) \right] = E_\theta^* \left[\nabla \cdot g(X, S) \right].$

2. *For any* $p \times p$ *matrix function* $T(X, S)$*, we have*

$$\begin{aligned} E_\theta \left[tr(T(X, S)) \Sigma^{-1} \right] = {} & 2C E_\theta^* \left[D_{1/2}^* T(X, S) \right] \\ & + C(n - p - 1) E_\theta^* \left[tr(S^{-1} T(X, S)) \right], \end{aligned}$$

where for any matrix A, $tr(A)$ is the trace of A, and where $D^*_{1/2}$ is the differential operator with respect to the variable S:

$$D^*_{1/2}g(X,S) = \sum_{i=1}^{p} \frac{\partial g_{ii}(X,S)}{\partial S_{ii}} + \frac{1}{2}\sum_{i\neq j} \frac{\partial g_{ij}(X,S)}{\partial S_{ij}}.$$

The shrinkage estimator introduced in Fourdrinier, Strawderman and Wells (2003) and Fourdrinier and Strawderman (2015) is $\delta(X,S) = X + g(X,S)$, where $g(\cdot)$ is some measurable function from $\mathbb{R}^p$ onto $\mathbb{R}^p$ and $S = VV^T$. We consider $g(X,S) = g(X)$ in (1.4). As a result, the soft-threshold shrinkage estimator can be the same as $\delta_\lambda^{soft}(X)$ in (1.3). In addition, we consider the target estimator $\delta_0(X)$ in (2.3). Under this condition, we have the following risk under BEL(δ_0) in (1.1):

$$\begin{aligned}
R_{\omega,\delta_0}(\theta,\delta_\lambda^{soft}(X)) &= E\left[L_{\omega,\delta_0}(\theta,\delta_\lambda^{soft}(X))\right] \\
&= E\Big(\omega\left(\delta_\lambda^{soft}(X)-\delta_0(X)\right)^T\Sigma^{-1}\left(\delta_\lambda^{soft}(X)-\delta_0(X)\right) \\
&\quad +(1-\omega)\left(\delta(X)-\theta\right)^T\Sigma^{-1}\left(\delta(X)-\theta\right)\Big) \\
&= E\Big(\omega^3 g^T(X)\Sigma^{-1}g(X) \\
&\quad +(1-\omega)\left(X+g(X)-\theta\right)^T\Sigma^{-1}\left(X+g(X)-\theta\right)\Big) \\
&= E\Big[(\omega^3-\omega+1)g^T(X)\Sigma^{-1}g(X) + 2g^T(X)\Sigma^{-1}(X-\theta) \\
&\quad +(1-\omega)\Big(\left(X-\theta\right)^T\Sigma^{-1}\left(X-\theta\right)\Big)\Big]. \qquad (2.7)
\end{aligned}$$

Under the above conditions and for $g(X_i)$ in (1.4), we have the following theorems.

Theorem 3. *Let (X,V) be an $n+1$ random vector in $\mathbb{R}^p$ with the elliptically contoured distribution in (2.5), with known Σ, and let $S = VV^T$. Assume that $g(X)$ is a function onto $\mathbb{R}^p$ that is weakly differentiable in X and differentiable in S. For the soft-threshold shrinkage estimator $\delta_\lambda^{soft}(X)$ and target estimator $\delta_0(X)$, the threshold λ^{sure} is given by*

$$\lambda^{sure} = \underset{0\le\lambda\le\lambda^U}{\operatorname{argmin}}\ SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X),\delta_0(X)),$$

where $SURE_{(X,\lambda)}(\delta_\lambda^{soft}(X),\delta_0(X))$ is

$$p(1-\omega) - 2\sum_{i=1}^{p} I(|X_i|\le\lambda)$$

$$+C(\omega^3-\omega+1)(n-p-1)E^*_\theta\left[\sum_{i=1}^{p} g^2(X_i)a_{ii}+\sum_{i\neq j} g(X_i)g(X_j)a_{ij}\right],$$

and $S^{-1}=A=(a_{ij})_{1\leq i,j\leq p}$.

Proof. Suppose that $T(X)=g(X)g^T(X)$. By Lemma 4 and Equation (2.7) under BEL(δ_0) in (1.1), $R_{\omega,\delta_0}(\theta,\delta^{soft}_\lambda(X))$ is equal to

$$\begin{aligned}&E\left[(1-\omega)\left(X-\theta\right)^T\Sigma^{-1}\left(X-\theta\right)\right]\\&+2C(\omega^3-\omega+1)E^*_\theta\left[D^*_{1/2}\left(g(X)g^T(X)\right)\right]\\&+C(\omega^3-\omega+1)(n-p-1)E^*_\theta\left[g^T(X)S^{-1}g(X)\right]+2E^*_\theta\left[\nabla\cdot g^T(X)\right].\end{aligned}$$

Suppose that $S^{-1}=A=(a_{ij})_{1\leq i,j\leq p}$. As a result $g^T(X)Ag(X)=\sum_{i=1}^p g^2(X_i)a_{ii}$ $+\sum_{i\neq j} g(X_i)g(X_j)a_{ij}$. Thus, the $SURE_{(X,\lambda)}$ is given as follows:

$$\begin{aligned}&p(1-\omega)-2\sum_{i=1}^{p} I(|X_i|\leq\lambda)\\&+C(\omega^3-\omega+1)(n-p-1)E^*_\theta\left[\sum_{i=1}^{p} g^2(X_i)a_{ii}+\sum_{i\neq j} g(X_i)g(X_j)a_{ij}\right].\end{aligned}$$

The proof is complete.

The value of the expectation $E^*_\theta[\sum_{i=1}^p g^2(X_i)a_{ii}+\sum_{i\neq j} g(X_i)g(X_j)a_{ij}]$ in Theorem 3 can be obtained using numerical methods. In addition, the multivariate normal distribution $X\sim N_p(\theta,\Sigma)$ is a special case of an elliptical distribution, in which case we have the following corollary, without computing $E^*_\theta(\cdot)$ using numerical methods.

Corollary 2. *Suppose the random variable* $X\sim N_p(\theta,\Sigma)$ *with known* Σ. *For the soft-threshold shrinkage estimator* $\delta^{soft}_\lambda(X)$ *and target estimator* $\delta_0(X)$, *the threshold* λ^{sure} *is given by*

$$\lambda^{sure}=\underset{0\leq\lambda\leq\lambda^U}{\operatorname{argmin}}\ SURE_{(X,\lambda)}(\delta^{soft}_\lambda(X),\delta_0(X)),$$

where $SURE_{(X,\lambda)}(\delta^{soft}_\lambda(X),\delta_0(X))$ *is given as follows:*

$$p(1-\omega)-2(1-\omega)\sum_{i=1}^{p} I(|X_i|<\lambda)+(\omega^3-\omega+1)\sum_{i=1}^{p} g^2(X_i)b_{ii}$$

$$+\sum_{i\neq j} g(X_i)g(X_j)b_{ij},$$

and $\Sigma^{-1} = B = (b_{ij})_{1\leq i,j\leq p}$.

Proof. The proof is similar to that of Theorem 3. However, we use Lemma 3 instead of Lemma 4.

3. SURE threshold of nonnegative location parameter

In this section, we discuss the SURE threshold in a nonnegative parameter space under a balanced loss function in the class of elliptical and spherical distributions. Mean vector (location) parameter estimation is an important problem in the context of shrinkage estimation, particularly when some components of the location parameter are restricted to a specific space.

As in Subsection 2.2, assume (X, U) is a $p+k$ random vector with a spherically symmetric distribution around the $p+k$ vector $(\theta, 0)$, dim $X =$ dim $\theta = p$, and dim $U =$ dim $0 = k$. Furthermore, suppose that the scalar variational component σ^2 is known, with density function (2.4). We wish to estimate $\theta = (\theta_1, \ldots, \theta_p)^T$ using $\delta_{\lambda}^{soft} = (\delta_1, \ldots, \delta_p)^T$ under a balanced loss function. Here, we consider cases where a subset of $\theta_i \geq 0$, for $i = 1, \ldots, p$, is nonnegative; that is, $\theta_1 \geq 0, \theta_2 \geq 0, \ldots, \theta_q \geq 0$ and $\theta_{q+1}, \theta_{q+2}, \ldots, \theta_p$ are unrestricted.

Define $\gamma_q(X) = (\gamma_q(X_1), \ldots, \gamma_q(X_p))$ as:

$$\gamma_q(X_j) = \begin{cases} -X_j, & X_j < 0, \\ 0, & X_j \geq 0, \end{cases} \quad j = 1, 2, \ldots, q \quad \text{and} \quad \gamma_q(X_j) = 0 \quad \text{if } j > q.$$

Then, similarly to Fourdrinier, Ouassou and Strawderman (2003) and Karamikabir, Afshari and Arashi (2018), the soft-threshold shrinkage estimators is defined as

$$\delta_{\lambda,q}^{soft}(X, U) = X + \gamma_q(X) + U^T U g(X). \tag{3.1}$$

In addition, we consider spherical distributions. Consider the following two target estimators:

$$\delta_0(X, U) = X + (1-\omega)U^T U g(X),$$
$$\delta_0^*(X) = X + (1-\omega)\gamma_q(X).$$

We can write $\delta_{\lambda,q}^{soft}(X, U) = \delta_0(X, U) + \gamma_q(X) + \omega U^T U g(X) = \delta_0^*(X) + \omega\gamma_q(X) + U^T U g(X)$. Note that all remarks in this section applied to cases in witch the subset of θ is nonnegative $(\theta_1 \geq 0, \theta_2 \geq 0, \ldots, \theta_q \geq 0)$.

Now, our goal is to find the risk for the soft-threshold shrinkage estimator $\delta^{soft}_{\lambda,q}(X,U)$ under BEL(δ_0) in (1.2). Thus, by using the target estimators $\delta_0(X,U)$ and $\delta^*_0(X)$ and by substituting $m=1$ in Lemma 2, we obtain the risks $R_{\omega,\delta_0(X,U)}(\theta,\delta^{soft}_{\lambda,q}(X,U))=R^1$ and $R_{\omega,\delta^*_0(X)}(\theta,\delta^{soft}_{\lambda,q}(X,U))=R^2$, as follows:

$$\begin{aligned}
R^1 &= \frac{1}{\sigma^2}E\left(\omega\|\delta^{soft}_{\lambda,q}(X,U)-\delta_0(X,U)\|^2+(1-\omega)\|\delta^{soft}_{\lambda,q}(X,U)-\theta\|^2\right)\\
&= \frac{1}{\sigma^2}E\Bigg[(\omega^3-\omega+1)\left(U^TU\right)^2\|g(X)\|^2+\|\gamma_q(X)\|^2\\
&\quad +2U^TU(\omega^2-\omega+1)\gamma^T_q(X)g(X)\\
&\quad +(1-\omega)\left(\|X-\theta\|^2+2\nabla.\gamma_q(X)+\frac{2}{k+2}\left(U^TU\right)^2\nabla.g(X)\right)\Bigg].\\
R^2 &= \frac{1}{\sigma^2}E\left(\omega\|\delta^{soft}_{\lambda,q}(X,U)-\delta^*_0(X)\|^2+(1-\omega)\|\delta^{soft}_{\lambda,q}(X,U)-\theta\|^2\right)\\
&= \frac{1}{\sigma^2}E\Bigg[\left(U^TU\right)^2\|g(X)\|^2+2U^TU(\omega^2-\omega+1)\gamma^T_q(X)g(X)\\
&\quad +(\omega^3-\omega+1)\|\gamma_q(X)\|^2\\
&\quad +(1-\omega)\left(\|X-\theta\|^2+2\nabla.\gamma_q(X)+\frac{2}{k+2}\left(U^TU\right)^2\nabla.g(X)\right)\Bigg].
\end{aligned}$$

In the following remark, similarly to Theorem 2, we obtain the $SURE$ using the above risk expressions.

Remark 1. Suppose that $(X,U)\sim SS_p((\theta,0),\sigma^2I_p)$ with known σ^2. For the soft-threshold shrinkage estimator $\delta^{soft}_{\lambda,q}(X,U)$ under the balanced loss BEL(δ_0) in (1.2), the threshold λ^{sure} is given by

$$\lambda^{sure}=\underset{0\le\lambda\le\lambda^U}{\operatorname{argmin}}\; SURE_{(X,\lambda)}(\delta^{soft}_{\lambda,q}(X,U),\delta_0),$$

where $SURE_{(X,\lambda)}(\delta^{soft}_{\lambda,q}(X,U),\delta_0)$ has the following cases:

1. For the target estimator $\delta_0(X,U)$, the risk estimate is equal to

$$\begin{aligned}
&\frac{(\omega^3-\omega+1)}{\sigma^2}\left(U^TU\right)^2\sum_{i=1}^p(|X_i|\wedge\lambda)^2\\
&+\frac{1}{\sigma^2}\sum_{i=1}^p X_i^2I(X_i<0)+p(1-\omega)\\
&+\frac{2(\omega^2-\omega+1)}{\sigma^2}\sum_{i=1}^p\gamma_q(X_i)g(X_i)\\
&+\frac{2(1-\omega)}{\sigma^2}\sum_{i=1}^pI(X_i<0)-\frac{2(1-\omega)}{\sigma^2(k+2)}\left(U^TU\right)^2\sum_{i=1}^pI(|X_i|\le\lambda).
\end{aligned}$$

2. For the target estimator $\delta_0^*(X)$, the risk estimate is equal to

$$\begin{aligned}&\frac{\left(U^TU\right)^2}{\sigma^2}\sum_{i=1}^{p}(|X_i|\wedge\lambda)^2+p(1-\omega)\\&+\frac{2(\omega^2-\omega+1)U^TU}{\sigma^2}\sum_{i=1}^{p}\gamma_q(X_i)g(X_i)\\&+\frac{(\omega^3-\omega+1)}{\sigma^2}\sum_{i=1}^{p}X_i^2I(X_i<0)+\frac{2(1-\omega)}{\sigma^2}\sum_{i=1}^{p}I(X_i<0)\\&-\frac{2(1-\omega)}{\sigma^2(k+2)}\left(U^TU\right)^2\sum_{i=1}^{p}I(|X_i|\le\lambda).\end{aligned}$$

As in Subsection 2.3, assume that $(X,V)=(X,V_1,\ldots,V_n)$ is an $n+1$ random vector in $\mathbb{R}^p$ with an elliptically contoured distribution of the form (2.5), with known Σ. We wish to estimate $\theta=(\theta_1,\ldots,\theta_p)^T$ using $\delta_\lambda^{soft}=(\delta_1,\ldots,\delta_p)^T$ under the balanced loss function and, again, we consider the cases where a subset of $\theta_i\ge 0$, $i=1,\ldots,p$ is nonnegative. The soft-threshold shrinkage and target estimators are respectively defined as follows:

$$\begin{aligned}\delta_{\lambda,q}^{soft}(X)&=X+\gamma_q(X)+g(X),\\\delta_0(X)&=X+(1-\omega)g(X).\end{aligned}$$

Again, our purpose is to find a risk for the soft-threshold shrinkage estimator $\delta_{\lambda,q}^{soft}(X)$ under BEL(δ_0) in (1.1). Suppose that $T(X)=g(X)g^T(X)$, $T^*(X)=\gamma(X)\gamma(X)^T$, $T^{**}(X)=\gamma(X)g^T(X)$, and $S^{-1}=A=(a_{ij})_{1\le i,j\le p}$. Using the target estimators $\delta_0(X)$ and $\delta_0^*(X)$ and Lemma 4, we obtain the risks $R_{\omega,\delta_0(X)}(\theta,\delta_{\lambda,q}^{soft}(X))=R^3$ and $R_{\omega,\delta_0^*(X)}(\theta,\delta_{\lambda,q}^{soft}(X))=R^4$, as follows:

$$\begin{aligned}R^3&=E\Big(\omega\left(\delta_{\lambda,q}^{soft}(X)-\delta_0(X)\right)^T\Sigma^{-1}\left(\delta_{\lambda,q}^{soft}(X)-\delta_0(X)\right)\\&\quad+(1-\omega)\left(\delta(X)-\theta\right)^T\Sigma^{-1}\left(\delta(X)-\theta\right)\Big)\\&=E\Big((1-\omega)(X-\theta)^T\Sigma^{-1}(X-\theta)+2(1-\omega)(X-\theta)^T\Sigma^{-1}\gamma_q(X)\\&\quad+2(1-\omega)(X-\theta)^T\Sigma^{-1}g(X)+2(\omega^2-\omega+1)\gamma_q^T(X)\Sigma^{-1}g(X)\\&\quad+(\omega^3-\omega+1)g^T(X)\Sigma^{-1}g(X)+\gamma_q^T(X)\Sigma^{-1}\gamma_q(X)\Big)\\&=E\Big((1-\omega)(X-\theta)^T\Sigma^{-1}(X-\theta)+2(1-\omega)\nabla\cdot\gamma_q(X)\\&\quad+2C(\omega^2-\omega+1)(n-p-1)E_\theta^*\left[\gamma_q(X)S^{-1}g(X)\right]\\&\quad+2(1-\omega)\nabla\cdot g(X)+4C(\omega^2-\omega+1)E_\theta^*\left[D_{1/2}^*\left(\gamma_q(X)g^T(X)\right)\right]\end{aligned}$$

$$+C(\omega^3-\omega+1)(n-p-1)E^*_\theta\left[g^T(X)S^{-1}g(X)\right]$$
$$+2C(\omega^3-\omega+1)E^*_\theta\left[D^*_{1/2}\left(g(X)g^T(X)\right)\right]$$
$$+C(n-p-1)E^*_\theta\left[\gamma_q^T(X)S^{-1}\gamma_q(X)\right]+2CE^*_\theta\left[D^*_{1/2}\big(\gamma_q(X)\gamma_q^T(X)\big)\right]\Big).$$

$$\begin{aligned}
R^4 &= E\Big(\omega\left(\delta^{soft}_{\lambda,q}(X)-\delta^*_0(X)\right)^T\Sigma^{-1}\left(\delta^{soft}_{\lambda,q}(X)-\delta^*_0(X)\right)\\
&\quad+(1-\omega)\left(\delta(X)-\theta\right)^T\Sigma^{-1}\left(\delta(X)-\theta\right)\Big)\\
&= E\Big((1-\omega)(X-\theta)^T\Sigma^{-1}(X-\theta)+2(1-\omega)(X-\theta)^T\Sigma^{-1}\gamma_q(X)\\
&\quad+2(1-\omega)(X-\theta)^T\Sigma^{-1}g(X)+2(\omega^2-\omega+1)\gamma_q^T(X)\Sigma^{-1}g(X)\\
&\quad+(\omega^3-\omega+1)\gamma_q^T(X)\Sigma^{-1}\gamma_q(X)+g^T(X)\Sigma^{-1}g(X)\Big)\\
&= E\Big((1-\omega)(X-\theta)^T\Sigma^{-1}(X-\theta)+2(1-\omega)\nabla\cdot\gamma_q(X)\\
&\quad+2(1-\omega)\nabla\cdot g(X)+2C(\omega^2-\omega+1)(n-p-1)E^*_\theta\left[\gamma_q(X)S^{-1}g(X)\right]\\
&\quad+4C(\omega^2-\omega+1)E^*_\theta\left[D^*_{1/2}\left(\gamma_q(X)g^T(X)\right)\right]\\
&\quad+C(\omega^3-\omega+1)(n-p-1)E^*_\theta\left[\gamma_q^T(X)S^{-1}\gamma_q(X)\right]\\
&\quad+2C(\omega^3-\omega+1)E^*_\theta\left[D^*_{1/2}\left(\gamma_q(X)\gamma_q^T(X)\right)\right]\\
&\quad+C(n-p-1)E^*_\theta\left[g^T(X)S^{-1}g(X)\right]+2CE^*_\theta\left[D^*_{1/2}\big(g(X)g^T(X)\big)\right]\Big).
\end{aligned}$$

In the following remark, similarly to Theorem 3, we obtain the $SURE$ using the above risks.

Remark 2. Let (X,V) an $n+1$ random vector in $\mathbb{R}^p$ following an elliptically contoured distribution (2.5) with known Σ and $S=VV^T$. Assume that $g(X)$ is a function on $\mathbb{R}^p$ that is weakly differentiable in X and differentiable in S. For the soft-threshold shrinkage estimator $\delta^{soft}_{\lambda,q}(X)$ under the balanced loss BEL(δ_0) in (1.1), the threshold λ^{sure} is given by

$$\lambda^{sure}=\underset{0\le\lambda\le\lambda^U}{\operatorname{argmin}}\ SURE_{(X,\lambda)}(\delta^{soft}_{\lambda,q}(X),\delta_0),$$

where $SURE_{(X,\lambda)}(\delta^{soft}_{\lambda,q}(X),\delta_0)$ has the following cases:

1. For the target estimator $\delta_0(X)$, the risk estimate is equal to

$$p(1-\omega)+2(1-\omega)I(X_i\le 0)-2(1-\omega)\sum_{i=1}^{p}I(|X_i|\le\lambda)$$
$$-2C(\omega^2-\omega+1)(n-p-1)(|X|\wedge\lambda)2E^*_\theta\left[\gamma_q(X)S^{-1}g(X)\right]$$
$$+C(\omega^3-\omega+1)(n-p-1)E^*_\theta\left[\sum_{i=1}^{p}g^2(X_i)a_{ii}+\sum_{i\ne j}g(X_i)g(X_j)a_{ij}\right]$$

$$+C(n-p-1)E_\theta^*\left[\sum_{i=1}^{p}\gamma_q^2(X_i)a_{ii}+\sum_{i\neq j}\gamma_q(X_i)\gamma_q(X_j)a_{ij}\right].$$

2. For the target estimator $\delta_0^*(X)$, the risk estimate is equal to

$$\begin{aligned}&p(1-\omega)+2(1-\omega)I([X_i\leq 0])-2(1-\omega)I(|X_i|\leq\lambda)\\&-2C(\omega^2-\omega+1)(n-p-1)2E_\theta^*\left[\gamma_q(X)S^{-1}g(X)\right]\\&+C(\omega^3-\omega+1)(n-p-1)E_\theta^*\left[\sum_{i=1}^{p}\gamma_q^2(X_i)a_{ii}+\sum_{i\neq j}\gamma_q(X_i)\gamma_q(X_j)a_{ij}\right]\\&+C(n-p-1)E_\theta^*\left[\sum_{i=1}^{p}g^2(X_i)a_{ii}+\sum_{i\neq j}g(X_i)g(X_j)a_{ij}\right].\end{aligned}$$

4. Simulation study

In this section, we compare our theoretical outcomes with numerical computations and simulations to investigate the performance of the soft-wavelet shrinkage estimator. All calculations in this section are performed using the **R** software package. For denoising or shrinkage coefficients, one of the most important concepts in wavelets and denoising is using thresholds. Shrinking of the empirical wavelet coefficients works best in problems where the underlying set of the true coefficients of f is sparse. The wavelet shrinkage method algorithm is as follows:

1. The discrete wavelet transform is derived from the noisy observations. In other words, let $Y_1,\ldots,Y_n$ be data observed from the model

$$Y_i=f(X_i)+\eta_i, \tag{4.1}$$

where $\{\eta_i\}$ is some noise, and $\{X_i\}$ is some set of points from the domain of $f(\cdot)$. Typically, n is an integer power of two. If $\mathbf{W}$ denotes the discrete wavelet transform matrix, then multiplying equation (4.1) by the orthogonal matrix $\mathbf{W}$ yields

$$X=\mathbf{W}Y=\mathbf{W}f+\mathbf{W}\eta=\theta+\epsilon.$$

Note that the observations are sampled from distribution f, but with some noise, which we went to remove. To achieve this aim, observations or noisy data are converted to wavelet coefficients.

2. Using the threshold value, the wavelet coefficients are divided into two

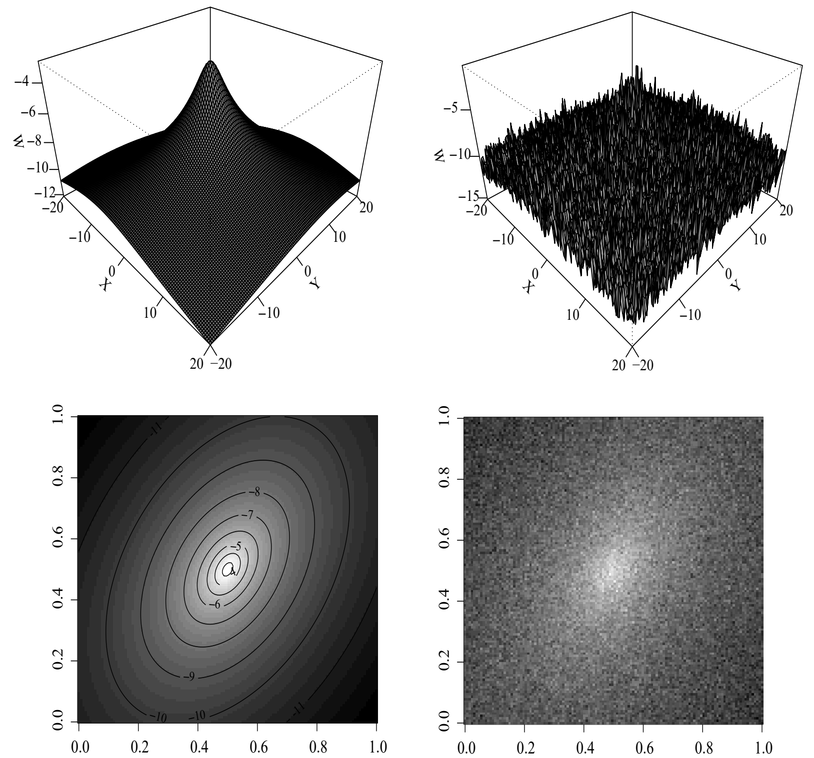

Figure 1. Top left: Density function of a $t_2(\mathbf{0}, \Sigma)$ with $N = 128$. Top right: Density function of a $t_2(\mathbf{0}, \Sigma)$ with added i.i.d. $N(0, 1)$ noise, with $N = 128$. Bottom left: Contour plot of a $t_2(\mathbf{0}, \Sigma)$ with $N = 128$. Bottom right: Contour plot of a $t_2(\mathbf{0}, \Sigma)$ with added i.i.d. $N(0, 1)$ noise, with $N = 128$.

groups: high effect (important) and low effect coefficients. If the wavelet coefficient is greater than the threshold value, it is categorized as an important coefficients, otherwise the set of coefficients negligible. Then, the low-effect coefficients are removed, and the important coefficients with respect to the shrinkage function are given by follows. In a hard shrinkage function, the coefficients are less than the threshold value of zero and the other coefficients remain unchanged. In a soft elimination function, the coefficients are less than the threshold value of zero, and the other coefficients decrease to the threshold value.

3. The last signal is reconstructed using the inverse wavelet transform.

For this purpose, a noise value using the standard normal distribution is added to the two-variate t-distribution $t_2(\mu, \Sigma)$, where $\mu = (0, 0)^T = \mathbf{0}$ and $\Sigma =$

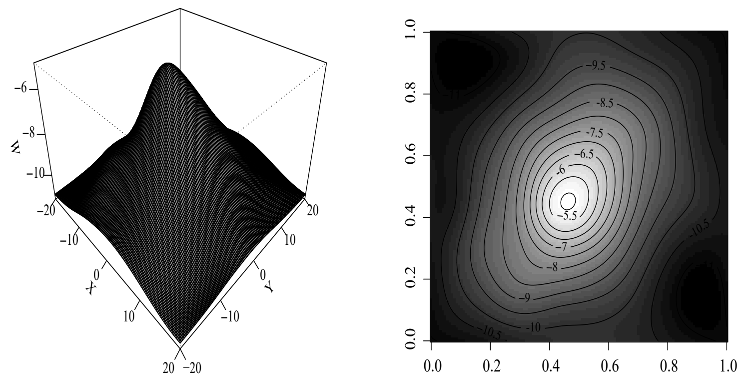

Figure 2. Density function and contour plot of a noisy $t_2(\mathbf{0}, \Sigma)$ signal after new threshold denoising with $N = 128$.

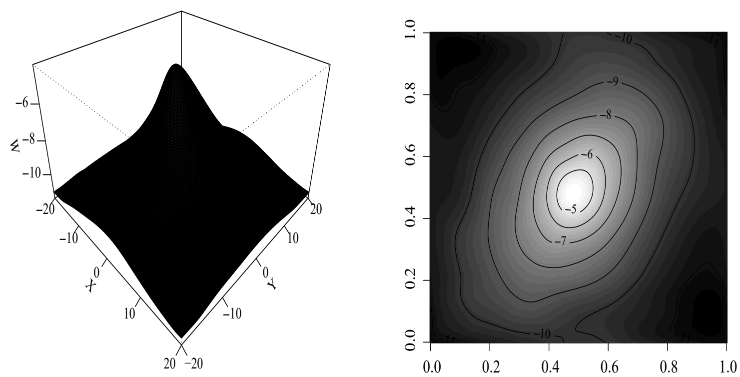

Figure 3. Density function and contour plot of a noisy $t_2(\mathbf{0}, \Sigma)$ signal after new threshold denoising with $N = 256$.

$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$. Figure 1 shows the density functions of $t_2(\mu, \Sigma)$ and $t_2(\mu, \Sigma)$ with added independent and identically distributed (i.i.d.) $N(0, 1)$ noise with $N = 128$ (noisy $t_2(\mathbf{0}, \Sigma)$) along with their contour plot. Now, by denoising the noisy $t_2(\mathbf{0}, \Sigma)$, we are looking for the main $t_2(\mathbf{0}, \Sigma)$. The minimization over the SURE is simply a grid search. The SURE value is first calculated for all X values, and then we find λ^{sure} from among these values. Using the SURE threshold, the noisy $t_2(\mathbf{0}, \Sigma)$ is converted to wavelet coefficients and denoised coefficients using the inverse discrete wavelet transformation, giving an approximation of the main $t_2(\mathbf{0}, \Sigma)$.

Figures 2, 3, and 4, for $N = 128, 256, 512$, show the density function and contour plot of the noisy $t_2(\mathbf{0}, \Sigma)$ signal after the SURE threshold denoising.

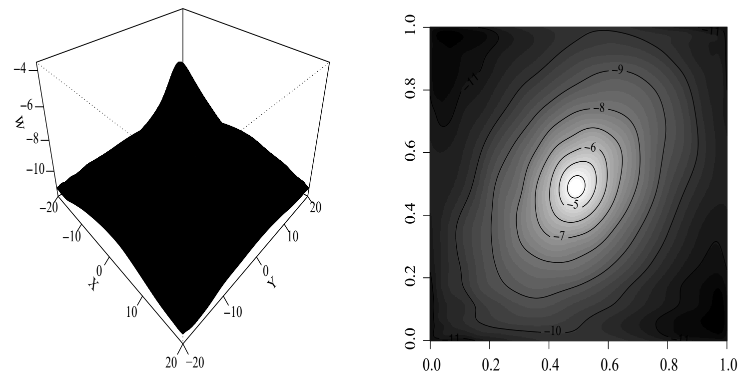

Figure 4. Density function and contour plot of a noisy $t_2(\mathbf{0}, \Sigma)$ signal after new threshold denoising with $N = 512$.

We compare the SURE method and four commonly used shrinkage strategies: hard and soft-thresholding with the universal threshold, CV, and Bayes thresholding. To assess the performance, we calculated the average mean squared error (AMSE) from $\mathbf{n} = 1{,}000$ replications of the simulation. The AMSE is obtained as follows:

$$\frac{1}{\mathbf{n}} \sum_{j=1}^{\mathbf{n}} \sum_{i=1}^{N} \frac{f(x_i) - \hat{f}(x_{i,j})}{N},$$

where $f(x_i)$ is the true signal, and $\hat{f}(x_{i,j})$ is the estimate of the function from simulation j. Lower values of AMSE represent a more accurate estimate.

Table 1, presents the AMSE with respect to p and σ^2 for the wavelet estimator of the target function based on the hard and soft universal threshold, CV, Bayes thresholding, and SURE for $\omega = 0.2$, $\omega = 0.5$ and $\omega = 0.8$. The simulations in this table assume an $N_2(\mathbf{0}, \sigma^2 I_2)$ distribution.

As can be seen in the table increasing σ^2 reduces the accuracy of all methods. In addition, the AMSE obtained using the SURE method is lower than that of the other methods, particularly when decreasing p. In general, the estimation accuracy decreases with an increase in ω.

We calculated the AMSE for the target estimators in the nonnegative location parameter space from 1,000 replications of the simulation. Table 2 shows the AMSE with respect to p, σ^2, and ω for the target estimators for $N_2(\mathbf{0}, \sigma^2 I_2)$. In this table, we can see the following:

- The AMSE of $\delta_0^{(1)}(X)$ for the corresponding values is lower than that of $\delta_0^{(2)}(X)$.

Table 1. AMSE for hard and soft universal thresholds, CV, Bayes thresholds, and SURE.

σ^2	p	SURE			Universal	Universal	CV	Bayes
		$\omega = 0.3$	$\omega = 0.5$	$\omega = 0.8$	Soft	Hard		threshold
	4	0.7303	0.7307	0.7326	0.7141	0.8542	–	0.7617
	8	0.5762	0.5781	0.5894	0.5805	0.7571	0.5797	0.6733
0.5	16	0.4948	0.5006	0.5248	0.5318	0.6581	0.5247	0.5997
	32	0.4511	0.4607	0.4870	0.5165	0.5761	0.5068	0.5464
	64	0.4293	0.4416	0.4715	0.5073	0.5410	0.4998	0.5213
	4	0.8756	0.8742	0.8707	0.8811	0.9642	–	0.9426
	8	0.7860	0.7765	0.7630	0.8682	0.9745	0.8422	0.9352
1	16	0.7394	0.7233	0.7038	0.9229	0.9939	0.8952	0.9584
	32	0.6968	0.6749	0.6514	0.9660	0.9879	0.9409	0.9772
	64	0.6668	0.6443	0.6247	0.9830	0.9954	0.9657	0.9885
	4	1.2249	1.2170	1.1991	1.2172	1.1801	–	1.3089
	8	1.2523	1.2212	1.1877	1.4512	1.4216	1.3699	1.4551
2	16	1.2364	1.2026	1.1440	1.7153	1.6719	1.6387	1.6986
	32	1.1730	1.1330	1.0618	1.8651	1.8133	1.8104	1.8420
	64	1.1241	1.0791	1.0131	1.9345	1.9032	1.8917	1.9210

Table 2. AMSE for target estimators.

		$\delta_0^{(1)}(X)$			$\delta_0^{(2)}(X)$		
σ^2	p	$\omega = 0.3$	$\omega = 0.5$	$\omega = 0.8$	$\omega = 0.3$	$\omega = 0.5$	$\omega = 0.8$
	4	0.7407843	0.7519258	0.8225013	0.7526715	0.7759163	0.8344153
	8	0.5957935	0.6122736	0.7792235	0.606251	0.6403176	0.8134959
0.5	16	0.4993819	0.5126939	0.5041723	0.5041723	0.5375324	0.8085041
	32	0.4561108	0.462669	0.7590374	0.4577497	0.4710982	0.8202242
	64	0.4452455	0.4394154	0.760832	0.4484006	0.4470308	0.831343
	4	0.8114424	0.7984361	0.8405045	0.8118768	0.8103646	0.8509157
	8	0.7162636	0.7110568	0.8133852	0.7180052	0.7234461	0.8324369
1	16	0.661435	0.6648597	0.8010723	0.6699519	0.6887543	0.8348478
	32	0.6443929	0.632851	0.7926348	0.6554189	0.6594202	0.8386649
	64	0.6331861	0.6178847	0.781318	0.6505679	0.645038	0.8312735
	4	0.9136572	0.8933877	0.896154	0.9043011	0.8834847	0.8984375
	8	0.8717649	0.8417345	0.8602557	0.8579672	0.8360454	0.8681166
2	16	0.8301697	0.7927979	0.8357576	0.8183334	0.794737	0.8558379
	32	0.8228262	0.7835409	0.8329445	0.8042471	0.7827455	0.8580542
	64	0.8224749	0.7801745	0.8206946	0.8023012	0.7719847	0.8491778

- Increasing σ^2 reduces the accuracy of all methods.
- Increasing the value of dim p increases the precision of the estimator.
- Increasing of the value of ω, decreases the estimation accuracy.

Figures 5 and 6 show the risk curves of the soft-threshold shrinkage estimator (3.1) for a nonnegative location parameter space and a multivariate normal distribution

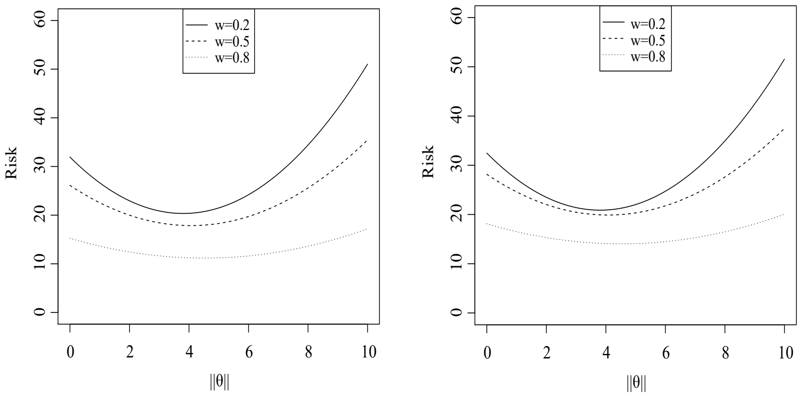

Figure 5. Risk curve for the target estimator $\delta_0(X, U)$ with $p = 16$, $q = 6$ (left), $q = 10$ (right), and for different values of ω.

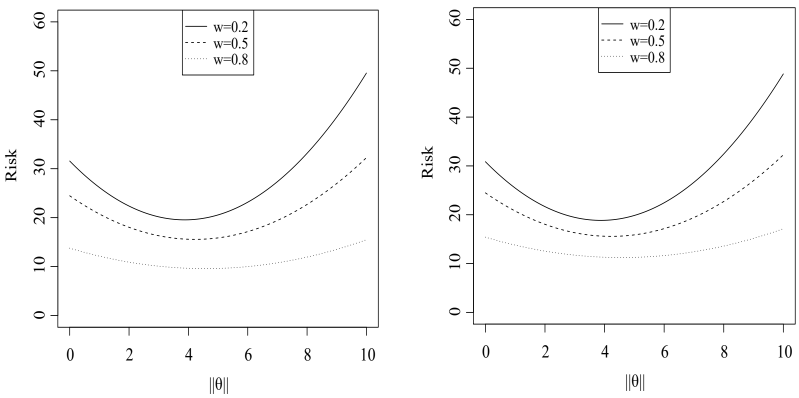

Figure 6. Risk curve for the target estimator $\delta_0^*(X)$ with $p = 16$, $q = 6$ (left), $q = 10$ (right), and for different values of ω.

$N_p(\mathbf{0}, 2I_p)$, with $p = 16$. These figures were plotted with respect to the target estimators $\delta_0(X, U)$, and $\delta_0^*(X)$ for $q = 6, 10$ and $\omega = 0.2, 0.5, 0.8$. Figures 5 and 6 also show that the risk is reduces by increasing ω. After increasing q from 6 to 10, the difference in the risk of the estimators decreases. In general, the risk values of the target estimator $\delta_0^*(X)$ are lower than those of $\delta_0(X, U)$.

5. Conclusion

We have generalized the SURE threshold for elliptical and spherical multivariate distributions under a balanced loss function. We also found the SURE threshold for the nonnegative mean vector of these distributions. The performance of the soft-threshold shrinkage estimator with SURE thresholding was investigated using a simulation study. The results demonstrate that the target estimator is appropriate, and that increasing the sample size, increases the accuracy of the new shrinkage estimator. In addition, the AMSE of the soft-threshold shrinkage estimator is shown to be lower than that of other estimators.

References

Afshari, M., Lak, F. and Gholizadeh, B. (2017). A new Bayesian wavelet thresholding estimator of nonparametric regression. *Journal of Applied Statistics* **44**, 649–666.

Antoniadis, A. (2007). Wavelet methods in statistics: Some recent developments and their applications. *Statistics Surveys* **1**, 16–55.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300.

Cai, T. T. and Silverman, B. W. (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya: The Indian Journal of Statistics* **63**, 127–148.

Cao, M. X. and He, D. (2017). Admissibility of linear estimators of the common mean parameter in general linear models under a balanced loss function. *Journal of Multivariate Analysis* **153**, 246–254.

Chang, Y. T. and Strawderman, W. E. (2017). Simultaneous estimation of p positive normal means with common unknown variance. *Statistics and Probability Letters* **121**, 83–89.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* **90**, 1200–1224.

Fourdrinier, D. and Strawderman, W. E. (1996). A paradox concerning shrinkage estimators: should a known scale parameter be replaced by an estimated value in the shrinkage factor. *Journal of Multivariate Analysis* **59**, 109–140.

Fourdrinier, D. and Ouassou, I. (2000). Estimation of the mean of a spherically symmetric distribution with constraints on the norm. *Canadian Journal of Statistics* **28**, 399–415.

Fourdrinier, D., Ouassou, I. and Strawderman, W. E. (2003). Estimation of a parameter vector when some components are restricted. *Journal of Multivariate Analysis* **86**, 14–27.

Fourdrinier, D., Strawderman, W. E. and Wells, M. T. (2003). Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix. *Journal of Multivariate Analysis* **85**, 24–39.

Fourdrinier, D. and Strawderman, W. E. (2015). Robust minimax Stein estimation under invariant data-based loss for spherically and elliptically symmetric distributions. *Metrika* **78**, 461–484.

Haff, L. T. R. (1979). An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis* **9**, 531–544.

Jafari Jozani, M., Marchand, E. and Parsian, A. (2006). On estimation with weighted balanced-type loss function. *Statistics and Probability Letters* **76**, 733–780.

Karamikabir, H., Afshari, M. and Arashi, M. (2018). Shrinkage estimation of non-negative mean vector with unknown covariance under balance loss. *Journal of Inequalities and Applications.* Doi:10.1186/s13660-018-1919-0.

Karamikabir, H. and Afshari, M. (2019). Wavelet shrinkage generalized Bayes estimation for elliptical distribution parameter's under LINEX loss. *International Journal of Wavelets, Multiresolution and Information Processing* **17**, 1950009. In press.

Krebs, J. T. N. (2018). Nonparametric density estimation for spatial data with wavelets. *Journal of Multivariate Analysis* **166**, 300–319.

Lorentz, M. (1905). Methods of measuring the concentration of wealth. *American Statistical Association* **9**, 209–219.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 674–693.

Nason G. P. (1996). Wavelet shrinkage using crossvalidation. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 463–479.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics* **9**, 1135–1151.

Van Eeden, C. (2006). *Restricted Parameter Space Estimation Problems, Admissibility and Minimaxity Properties.* Springer, New York.

Vidakovic, B. (2009). *Statistical Modelling by Wavelets.* John Wiley and Sons, New York.

Zellner, A. (1994). Bayesian and non-Bayesian estimation using balanced loss functions. In *Statistical Decision Theory and Methods* (Edited by J. O. Berger and S. S. Gupta), 337–390. Springer, New York.

Hamid Karamikabir

Department of Statistics, Persian Gulf University, Bushehr, Iran.

E-mail: h_karamikabir@yahoo.com

Mahmoud Afshari

Department of Statistics, Persian Gulf University, Bushehr, Iran.

E-mail: afshar@pgu.ac.ir

(Received January 2019; accepted January 2020)