<table>
<thead>
<tr>
<th>Statistica Sinica Preprint No: SS-2019-0339</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
New Wavelet SURE Thresholds of Elliptical Distributions
under the Balance Loss

Hamid Karamikabir and Mahmoud Afshari

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

Abstract: In this paper, a new shrinkage soft wavelet threshold estimator based on Stein’s unbiased risk estimators (SURE) is introduced for elliptical and spherical distributions under the balance loss function. Our focus will be on a particular thresholding rules to obtain a new threshold to produce new estimators. Also, we obtain SURE shrinkage based on non-negative subset of mean vector. Finally, we present a simulation in order to test the validity of the purposed estimator.

Key words and phrases: Balance loss function, Restricted estimator, Shrinkage estimator, Spherically distribution, Elliptically distribution, Soft wavelet estimator.

Corresponding author. Email: afshar@pgu.ac.ir
1. Introduction

Mean vector (location) parameter estimation is an important problem in the context of the point estimation. At point estimation, we want to obtain an estimator that minimizes the risk function for all possible parameter values. But in practice, given the size of the class of estimators, this is not possible. For evaluating the performance of estimators, we need to set a measure. Location vector parameter estimation is done through a variety of loss functions. One of the types of loss functions is the balanced loss function. In this paper, we use the balanced loss function in the following two ways.

Definition 1. Suppose that $X$ is a random vector with mean vector parameter $\theta$ and scalar variational component $\Sigma$. The balanced error loss function, $\text{BEL}(\delta_0)$, is defined as

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

where $0 \leq \omega < 1$, $\delta_0$ ia a target estimator and $\Sigma$ is symmetric non-singular scale matrix.

As a special case, suppose that $X$ is a random vector with mean vector parameter $\theta$ and scalar variational component $\sigma^2$ (i.e. $\Sigma = \sigma^2 I_p$, in
Definition (1]). Then the balanced error loss function is as follows:

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

The corresponding risk is the expectation with respect to the loss function. Then the associated risk function with respect to (1.2) will be \( R(\theta, \delta) = E_\theta [L(\theta, \delta)] \). 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 more details about the use of this loss, we refer to Jafari Jozani et al. (2006), Cao and He (2017) and Karamikabir et al. (2018).

Shrinking and truncating the data directly or the coefficients in their Fourier series expansions is an old technique in signal processing. For non-local bases, such as the trigonometric, shrinking the coefficients can affect the global shape of the reconstructed function and introduce unwanted artifacts. In the context of function estimation by wavelets, the shrinkage has an additional feature; it is connected with smoothing (de-noising) because the measure of smoothness of a function depend on the magnitudes of its wavelet coefficients (Vidakovic (2009)).

Shrinkage estimation is a method to improve 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 minimum quadratic risk com-
pared to natural estimators (mostly the maximum likelihood estimator). The general form of the shrinkage estimator is $X + g(X)$. The shrinkage is usually done by decreasing amount of $X$ 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 of selecting a threshold by minimizing Stein’s unbiased estimator of risk (Stein, 1981). This threshold is implemented in an adapting de-noising procedure, SureShrink. The adaptation in SureShrink is achieved by specifying thresholds level-wise.

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. Soft thresholding wavelet shrinkage estimation by Antoniadis (2007) is given by the following:

$$
\delta^\text{soft}_\lambda(X_i) = (X_i - \text{sign}(X_i)\lambda)I(|X_i| > \lambda) = \begin{cases} 
X_i + \lambda & X_i < -\lambda, \\
0 & |X_i| \leq \lambda, \\
X_i - \lambda & X_i > \lambda,
\end{cases} \quad (1.3)
$$

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

$$\delta_{\lambda}^{soft}(X) = X + g(X), \quad (1.4)$$

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

$$g(X_i) = \begin{cases} 
\lambda & X_i < -\lambda, \\
-X_i & |X_i| \leq \lambda, \\
-\lambda & X_i > \lambda. 
\end{cases} \quad (1.5)$$

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

In this paper, we generalize Donoho and Johnstone (1995) results for Stein’s unbiased risk estimators (SURE) thresholds with changes in the class of family distributions and loss function. For this purpose, we derive new estimators of the location parameter of the multivariate normal, elliptical or spherically symmetric distribution in classes of shrinkage estimators that are well behaved under the balanced error loss. These results generalize previous work by Donoho and Johnstone (1995), Fourdrimer and
New Wavelet SURE Thresholds of Elliptical Distributions

Ouassou (2000), Fourdrinier et al. (2003a) and Fourdrinier and Strawderman (2015) with the shrinkage wavelet approach. We derive estimates of the risk using Stein’s lemma of the risk of those estimators, and derive expressions for the optimal \( \lambda \) under the SURE in the multivariate normal, spherically symmetric and elliptical case. Also, we find the SURE shrinkage in the restricted parameter space. In fact, we generalized previous work by Fourdrinier et al. (2003a) and Karamikabir et al. (2018) with the shrinkage wavelet approach. In this respect, Chang and Strawderman (2017) studied a shrinkage estimation of \( p \) positive normal means under sum of squared errors loss. Also, recently Afshari et al. (2017), Krebs (2018) and Karamikabir and Afshari (2019) have been working on the application of wavelets for multivariate distributions. For more details on this topic, we refer to Van Eeden (2006) and Karamikabir et al. (2018), among others.

Shrinkage wavelet has many applications in statistical estimation. Some applications of shrinkage wavelet lie on the cross-boundaries between disciplines. For instance, recent applications of wavelets in the shape analysis are bridging the areas of statistical modeling, statistical theory of shapes, computational geometry, and image processing. For more information see Vidakovic (2009).

Parameter estimation is important in statistical inference, specially
when the goal is to estimate the parameter vector in a multivariate distribution. A well-known class of multivariate distributions is the 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 $\theta$, $p \times p$ scale matrix $\Sigma$ and the characteristic generator $\psi$, if its density function is of the form

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

(1.6)

where $C_m$ is normalizing constant. Its characteristic function satisfies

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

Some examples of elliptical distributions are the multivariate normal distribution $N_p(\theta, \Sigma)$, the multiuniform distribution, the multivariate Pearson type II and VII distributions, the multivariate Laplace distribution, the generalized slash distribution, the multivariate Cauchy distribution, the multivariate Bessel distribution, the multivariate exponential power distribution and the 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 $\Lambda X$ have the same distribution for all $p \times p$ orthogonal matrices $\Lambda$. The elliptical family are
the 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 identity matrix. Some
examples of spherical distributions are the multivariate normal distribution $N_p(0, \sigma^2 I_p)$, the $\varepsilon$-contaminated normal distribution, the multivariate
$t$ distribution and the scale mixture of multivariate normal distributions.

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

The paper is divided as follows: in Section 2, we find a threshold based
on SURE under the balanced loss function in the class of elliptical and
spherical distributions. Section 3, we discuss the SURE shrinkage in the
restricted parameter space, while the numerical performance analysis is
investigated by a simulation study in Section 4. Section 5 presents the
conclusion of the paper.

2. A Threshold Based on SURE

In this section, we first review the method of Donoho and Johnstone (1995)
for finding the SURE threshold in a multivariate normal distribution $N_p(\theta, I_p)$. 

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)
In the next step, we try to 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 diagonal scale matrix $\sigma^2 I_p$, i.e. $N_p(\theta, \sigma^2 I_p)$, a spherically symmetric distribution $SS(\theta, \sigma^2 I_p, \psi)$ and finally an elliptical distribution $E_p(\theta, \Sigma, \psi)$.

For all these, instead of 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 $\sigma^2$, then

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

where $g : \mathbb{R}^p \rightarrow \mathbb{R}^p$ 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 Stein’s unbiased estimator of risk. This threshold is implemented in an adaptive de-noising procedure, SureShrink. The adaptation in SureShrink is achieved by level-wise specification. In detail, suppose that $\theta = (\theta_1, \ldots, \theta_p)^T$ and $X_i \sim N(\theta_i, 1), i = 1, \ldots, p$. Let $\delta(X)$ be
an estimator of $\theta$. If the function $g = \{g_i\}_{i=1}^p$ in the shrinkage estimator representation $\delta(X) = X + g(X)$ is weakly differentiable, (that is, if there exists a function $q_\theta(x)$ such that $\partial/\partial \theta E_\theta[\delta(X)] = E[\delta(q_\theta))]$) then under quadratic loss function $L(\theta, \delta) = \|\delta(X) - \theta\|^2$ by Lemma 1 we have the following expectation.

$$R(\theta, \delta(X)) = E_\theta \left(\|\delta(X) - \theta\|^2\right) = p + E_\theta \left(\|g(X)\|^2\right) + E_\theta (2 \nabla \cdot g(X))(2.7)$$

We consider the shrinkage soft threshold estimator $\delta^{soft}_\lambda(X) = X + g(X)$ in (1.4) where $g(X)$ is in (1.5). We need the following useful equations for $g(X)$.

$$\nabla \cdot g(X) = -\sum_{i=1}^P I(|X_i| \leq \lambda), \quad \|g(X)\|^2 = \sum_{i=1}^P (|X_i| \wedge \lambda)^2. \quad (2.8)$$

where $(|X_i| \wedge \lambda) = \min(|X_i|, \lambda)$. Since $\delta^{soft}_\lambda(X)$ is weakly differentiable in Stein’s sense and by using Lemma 1 and Equation (2.8), we get from (2.7) that the quantity

$$SURE(X, \lambda) = p - 2 \sum_{i=1}^P I(|X_i| \leq \lambda) + \sum_{i=1}^P (|X_i| \wedge \lambda)^2 \quad (2.9)$$

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

$$\lambda^{sure} = \arg \min_{0 \leq \lambda \leq \lambda^u} SURE(X, \lambda). \quad (2.10)$$
Arguing heuristically, one expects that, for large dimension $X$, a sort of statistical regularity will set in, the Law of Large Numbers will ensure that SURE is close to the true risk, and that $\lambda^{SURE}$ will be almost the optimal threshold for the case at hand.

The $\lambda^{sure}$ 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 diagonal scale matrix

Now, under above condition, we consider the balanced loss function (1.2). In this case similar to Jafari Jozani et al. (2006) and Karamikabir et al. (2018), the target estimator can be the part of shrinkage soft threshold estimator $\delta^{soft}_\lambda(X)$. The target estimator is as follows.

$$\delta_0(X) = X + (1 - \omega)g(X).$$  \hspace{1cm} (2.11)

And hence the shrinkage soft threshold estimator is $\delta^{soft}_\lambda(X) = \delta_0(X) + \omega g(X)$.

**Theorem 1.** Suppose that $X \sim N_p(\theta, \sigma^2 I_p)$ with known $\sigma^2$. For the shrinkage soft threshold estimator $\delta^{soft}_\lambda(X)$ and target estimator $\delta_0(X)$, the value
2.1 Multivariate normal distribution with diagonal scale matrix

of threshold is given by

$$
\lambda^{\text{sure}} = \arg \min_{0 \leq \lambda \leq \lambda^U} SURE_{(X, \lambda)}(\delta^{\text{soft}}_\lambda(X), \delta_0(X)),
$$

where $SURE_{(X, \lambda)}(\delta^{\text{soft}}_\lambda(X), \delta_0(X))$ is 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 4 and under $BEL(\delta_0)$ in (1.2), the $R_{\omega, \delta_0}(\theta, \delta^{\text{soft}}_\lambda(X))$ is as follows.

$$
E \left[ L_{\omega, \delta_0}(\theta, \delta^{\text{soft}}_\lambda(X)) \right] = \frac{1}{\sigma^2} E \left( \omega \| \delta^{\text{soft}}_\lambda(X) - \delta_0(X) \|^2 + (1 - \omega) \| \delta^{\text{soft}}_\lambda(X) - \theta \|^2 \right)
$$

$$
= \frac{1}{\sigma^2} E \left( \omega \| X + g(X) - X - (1 - \omega) g(X) \|^2 + (1 - \omega) \| X + g(X) - \theta \|^2 \right)
$$

$$
= \frac{1}{\sigma^2} E \left( \omega^3 \| g(X) \|^2 + (1 - \omega) \| X - \theta \|^2 + (1 - \omega) \| g(X) \|^2 + 2(1 - \omega)(X - \theta)^T g(X) \right)
$$

$$
= \frac{1}{\sigma^2} E \left( \omega^3 - \omega + 1 \| g(X) \|^2 + (1 - \omega) \| X - \theta \|^2 + 2(1 - \omega) \nabla \cdot g(X) \right).
$$

Then the $SURE_{(X, \lambda)}(\delta^{\text{soft}}_\lambda(X), \delta_0(X))$ is equal to

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

The $SURE_{(X, \lambda)}$ is an unbiased estimate of risk, i.e.,

$$
E \left[ L_{\omega, \delta_0}(\theta, \delta^{\text{soft}}_\lambda(X)) \right] = E \left[ SURE_{(X, \lambda)}(\delta^{\text{soft}}_\lambda(X), \delta_0(X)) \right],
$$
and $\lambda_{\text{sure}} = \arg \min_{0 \leq \lambda \leq \lambda_U} SURE(X, \lambda)(\delta^{\text{soft}}(X), \delta_0(X))$. 

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

$$f_{X,U}(x, u) = \frac{1}{\sigma^{p+k}} f\left(\frac{\|x - \theta\|^2 + u^T u}{\sigma^2}\right), \quad (2.12)$$

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

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

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

Let the target estimator be
\[ \delta_0(X, U) = X + (1 - \omega) U^T g(X), \]  
and hence we can write \( \delta_{\lambda}^{soft}(X, U) = \delta_0(X, U) + \omega U^T g(X). \)

**Theorem 2.** Suppose that \((X, U) \sim SS_p((\theta, 0), \sigma^2 I_p)\) with known \( \sigma^2 \). For the shrinkage soft threshold estimator \( \delta_{\lambda}^{soft}(X, U) \) and target estimator \( \delta_0(X, U) \), threshold \( \lambda^{sure} \) is given by
\[ \lambda^{sure} = \arg \min_{0 \leq \lambda \leq \lambda^U} SURE_{(X, \lambda)}(\delta_{\lambda}^{soft}(X, U), \delta_0(X, U)). \]

where \( SURE_{(X, \lambda)}(\delta_{\lambda}^{soft}(X, U), \delta_0(X, U)) \) is as follows.
\[ \frac{(\omega^3 - \omega + 1)}{\sigma^2} \sum_{i=1}^{p} (|X_i| \wedge \lambda)^2 \]
\[ + p(1 - \omega) \frac{2(1 - \omega)}{\sigma^2(k + 2)} (U^T U)^2 \sum_{i=1}^{p} I(|X_i| \leq \lambda). \]

**Proof.** By replacing \( m = 1 \) in Lemma 2 we have the following risk.
\[ R_{\omega, \delta_0}(\theta, \delta_{\lambda}^{soft}) = E \left[ L_{\omega, \delta_0}(\theta, \delta_{\lambda}^{soft}(X, U)) \right] \]
2.3 Elliptical distribution

\[
\frac{1}{\sigma^2} E \left( \omega \| \delta^\text{soft}_\lambda (X, U) - \delta_0 (X, U) \|^2 
+ (1 - \omega) \| \delta^\text{soft}_\lambda (X, U) - \theta \|^2 \right) 
= \frac{1}{\sigma^2} E \left( \omega \| X + U^T g(X) - X - (1 - \omega) U^T U g(X) \|^2 
+ (1 - \omega) \| X + U^T U g(X) - \theta \|^2 \right) 
= \frac{1}{\sigma^2} E \left[ \omega^3 (U^T U)^2 \| g(X) \|^2 
+ (1 - \omega) \left( \| X - \theta \|^2 + (U^T U)^2 \| g(X) \|^2 
+ 2(X - \theta)^T U^T U g(X) \right) \right] 
= \frac{1}{\sigma^2} E \left[ (\omega^3 - \omega + 1) (U^T U)^2 \| g(X) \|^2 
+ (1 - \omega) \left( \| X - \theta \|^2 + \frac{2}{k+2} (U^T U)^2 \nabla \cdot g(X) \right) \right].
\]

So, \( SURE_{(X, \lambda)}(\delta^\text{soft}_\lambda (X, U), \delta_0 (X, U)) \) is as follows.

\[
\frac{1}{\sigma^2} \frac{(\omega^3 - \omega + 1) (U^T U)^2}{\sigma^2} \sum_{i=1}^{p} (|X_i| \wedge \lambda)^2 
+ p(1 - \omega) - \frac{2(1 - \omega)}{\sigma^2(k+2)} (U^T U)^2 \sum_{i=1}^{p} I(|X_i| \leq \lambda).
\]

Also, the \( SURE_{(X, \lambda)} \) is an unbiased estimate of risk, i.e.,

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

and \( \lambda^\text{sure} = \arg \min_{0 \leq \lambda \leq \lambda^U} SURE_{(X, \lambda)}(\delta^\text{soft}_\lambda (X, U), \delta_0 (X, U)). \)

2.3 Elliptical distribution

In this subsection, the statistical distribution is in the form of the elliptical contoured distribution in [Fourdrinier et al. (2003a)].
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
\[
f(x, v) = |\Sigma^{-1}| f \left( (x - \theta)^T \Sigma^{-1} (x - \theta) + \sum_{i=1}^{n} v_i^T \Sigma^{-1} v_i \right), \tag{2.15}
\]
where \(X\) and the \(V_i\) are \(p \times 1\) vectors, \(\theta\) is \(p \times 1\) unknown location vector and \(\Sigma\) is \(p \times p\) known matrix proportional to the covariance matrix. This density arises as a joint density in the canonical form of the general linear model, where \(X\) is a projection on the space spanned by \(\theta\) and \(V_i\)'s are projections onto the orthogonal complement of the space spanned by \(\theta\) or \(X \sim E(\theta, \Sigma, \psi)\) and \(V = (V_1, \ldots, V_n) \sim E(0, \Sigma, \psi)\). The class in (2.15) contains models such as the multivariate normal, \(t\) and Kotz-type distributions.

In the case of normal distribution, that is, \(X \sim N_p(\theta, \Sigma), V = (V_1, \ldots, V_n)\) where the \(V_i \sim N_p(0, \Sigma)\) are independent columns vectors and \(S = V V^T \sim W_p(n, \Sigma)\) (Wishart distribution) with \(n \geq p\), the following lemma is a straightforward extension of Stein \((1981)\)'s lemma and of Haff \((1979)\)'s lemma.

**Lemma 3.** \((\text{Fourdrinier and Strawderman} \ (2013))\) Assume that \((X, V) = (X, V_1, \ldots, V_n)\) be an \(n + 1\) random vector in \(\mathbb{R}^p\) with a multivariate normal distribution where \(X \sim N_p(\theta, \Sigma), V = (V_1, \ldots, V_n), V_i \sim N_p(0, \Sigma)\).
Also let $g(X, S)$ be a $p \times 1$ vector such that the function $g(X, \cdot)$ is weakly differentiable, then we have

$$E_\theta \left[ (X - \theta)^T \Sigma^{-1} g(X, S) \right] = E_\theta \left[ \nabla \cdot g(X, S) \right], \quad (2.16)$$

provided the expectations (2.16) exist. As defined before, $\nabla \cdot g(X, S)$ is the divergence operator with respect to the variable $X$.

Similar to Fourdrinier et al. (2003b), we define expectations $E^*_\theta$ with respect to the distribution $C^{-1} F \left( (X - \theta)^T \Sigma^{-1} (X - \theta) + \sum_{j=1}^n V_j^T \Sigma^{-1} V_j \right)$, where $F$ and $C$ are defined as

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

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

**Lemma 4.** (Fourdrinier et al. (2003b)) Let $(X, V)$ be $n + 1$ random vector in $\mathbb{R}^p$ with elliptically contoured distribution and let $S = VV^T$. Assume that $g(X, S)$ is a function into $\mathbb{R}^p$ which 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

$$E_\theta \left[ tr(T(X, S)) \Sigma^{-1} \right] = 2CE^*_\theta \left[ D_{1/2}^* T(X, S) \right]$$

$$+ C(n - p - 1) E^*_\theta \left[ tr(S^{-1} T(X, S)) \right]. \quad (2.17)$$
2.3 Elliptical distribution

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

$$D_{1/2}^*(X, S) = \sum_{i=1}^p \frac{\partial g_{ii}(X, S)}{\partial S_{ii}} + \frac{1}{2} \sum_{i \neq j} \frac{\partial g_{ij}(X, S)}{\partial S_{ij}}. \quad (2.18)$$

The shrinkage estimator introduced in Fourdrinier et al. (2003b) and Fourdrinier and Strawderman (2013) is $\delta(X, S) = X + g(X, S)$ where $g(\cdot)$ is some measurable function from $\mathbb{R}^p$ into $\mathbb{R}^p$ and $S = VV^T$. We consider $g(X, S) = g(X)$ in (1.5). As a result, the shrinkage soft threshold estimator can be the same as $\delta_\lambda^{soft}(X)$ in (1.4). Also, we consider the target estimator equal to $\delta_0(X)$ in (2.11). Under this condition, we have the following risk under $\text{BEL}(\delta_0)$ in (1.1):

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

$$= E \left( \omega \left( \delta_\lambda^{soft}(X) - \delta_0(X) \right) ^T \Sigma^{-1} \left( \delta_\lambda^{soft}(X) - \delta_0(X) \right) + (1 - \omega) (\delta(X) - \theta)^T \Sigma^{-1} (\delta(X) - \theta) \right)$$

$$= E \left( \omega^3 g^T(X) \Sigma^{-1} g(X) + (1 - \omega) (X + g(X) - \theta)^T \Sigma^{-1} (X + g(X) - \theta) \right)$$

$$= E \left[ (\omega^3 - \omega + 1)g^T(X) \Sigma^{-1} g(X) + 2g^T(X) \Sigma^{-1} (X - \theta) + (1 - \omega) \left( (X - \theta)^T \Sigma^{-1} (X - \theta) \right) \right]. \quad (2.19)$$

Under the above conditions and for $g(X_i)$ in (1.5), we have the following theorems.
2.3 Elliptical distribution

**Theorem 3.** Let \((X, V)\) be \(n + 1\) random vector in \(\mathbb{R}^p\) with elliptically contoured distribution (2.15) with known \(\Sigma\) and let \(S = VV^T\). Assume that \(g(X)\) is a function into \(\mathbb{R}^p\) which is weakly differentiable in \(X\) and differentiable in \(S\). For the shrinkage soft threshold estimator \(\delta^\text{soft}_\lambda(X)\) and target estimator \(\delta_0(X)\), threshold \(\lambda^{\text{sure}}\) is given by

\[
\lambda^{\text{sure}} = \arg \min_{0 \leq \lambda \leq \lambda^U} \text{SURE}_{(X, \lambda)}(\delta^\text{soft}_\lambda(X), \delta_0(X)),
\]

where the \(\text{SURE}_{(X, \lambda)}(\delta^\text{soft}_\lambda(X), \delta_0(X))\) is

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

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

**Proof.** Suppose that \(T(X) = g(X)g^T(X)\). By Lemma 1 and Equation (2.19) under \(\text{BEL}(\delta_0)\) in (1.1), the \(R_{\omega, \delta_0}(\theta, \delta^\text{soft}_\lambda(X))\) is equal to

\[
E \left[ (1 - \omega)(X - \theta)^T \Sigma^{-1} (X - \theta) \right]
+ 2C(\omega^3 - \omega + 1)E^*_\theta [D^*_{1/2} (g(X)g^T(X))] \\
+ C'(\omega^3 - \omega + 1)(n - p - 1)E^*_\theta [g^T(X)S^{-1}g(X)] + 2E^*_\theta [\nabla \cdot g^T(X)].
\]

Suppose that \(S^{-1} = A = (a_{ij})_{1 \leq i,j \leq p}\), as a result

\[
g^T(X)Ag(X) = \sum_{i=1}^p g^2(X_i)a_{ii} + \sum_{i \neq j} g(X_i)g(X_j)a_{ij}.
\]

So, the \(\text{SURE}_{(X, \lambda)}\)
2.3 Elliptical distribution

is as follows.

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

The proof is complete. \(\square\)

The value of expectation \(E_\theta \left[ \sum_{i=1}^{p} g^2(X_i)a_{ii} + \sum_{i \neq j} g(X_i)g(X_j)a_{ij} \right]\) 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 \(E_\theta^* (\cdot)\) by numerical methods.

**Corollary 2.** Suppose that random variable \(X \sim N_p(\theta, \Sigma)\) with known \(\Sigma\).

For the shrinkage soft threshold estimator \(\delta^{soft}_\lambda(X)\) and target estimator \(\delta_0(X)\), the threshold \(\lambda^{\text{sure}}\) is given by

\[
\lambda^{\text{sure}} = \arg \min_{0 \leq \lambda \leq c} SURE_{(X, \lambda)}(\delta^{soft}_\lambda(X), \delta_0(X)) ,
\]

where the \(SURE_{(X, \lambda)}(\delta^{soft}_\lambda(X), \delta_0(X))\) is 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 non-negative location parameter

In this section, we discuss the SURE threshold in the non-negative parameter space under the 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, specially when some components of location parameter are restricted to be situated in a specific space.

As in Subsection 2.2, assume \((X, U)\) is a \(p + k\) random vector having a spherically symmetric distribution around the \(p + k\) vector \((\theta, 0)\), \(\dim X = \dim \theta = p\) and \(\dim U = \dim 0 = k\). Further, suppose that the scalar variational component \(\sigma^2\) is known with density function (2.12). We wish to estimate \(\theta = (\theta_1, \ldots, \theta_p)^T\) by \(\delta_{\lambda}^{soft} = (\delta_1, \ldots, \delta_p)^T\) under the balanced loss function. Here, we consider the cases where a subset of \(\theta_i \geq 0, i = 1, \ldots, p\) are non-negative, i.e., \(\theta_1 \geq 0, \theta_2 \geq 0, \ldots, \theta_q \geq 0\) and that \(\theta_{q+1}, \theta_{q+2}, \ldots, \theta_p\) are unrestricted.
Define $\gamma_q(X) = (\gamma_q(X_1), \ldots, \gamma_q(X_p))$ as:

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

Then, similar to Fourdrinier et al. (2003a) and Karamikabir et al. (2018), the shrinkage soft threshold estimators are respectively defined as

$$
\delta_{\lambda_q}^{soft}(X, U) = X + \gamma_q(X) + U^T U g(X). \quad (3.20)
$$

Also, we consider spherical distributions. Consider the two following 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 $\delta_{\lambda_q}^{soft}(X, U) = \delta_0(X, U) + \gamma_q(X) + \omega U^T U g(X) = \delta_0^*(X) + \omega \gamma_q(X) + U^T U g(X)$. All the remarks of this section are in cases where the subset of $\theta$ is non-negative ($\theta_1 \geq 0, \theta_2 \geq 0, \ldots, \theta_q \geq 0$).

Now, our goal is to find the risk for the shrinkage soft threshold estimator $\delta_{\lambda_q}^{soft}(X, U)$ under BEL($\delta_0$) in (1.2). So, by using the target estimators $\delta_0(X, U)$ and $\delta_0^*(X)$ and by replacing $m = 1$ in Lemma 2, we obtain the risks $R_{\omega, \delta_0(X, U)}(\theta, \delta_{\lambda_q}^{soft}(X, U)) = R^1$ and $R_{\omega, \delta_0^*(X)}(\theta, \delta_{\lambda_q}^{soft}(X, U)) = R^2$ as follows.

$$
R^1 = \frac{1}{\sigma^2} E \left( \omega \|\delta_{\lambda_q}^{soft}(X, U) - \delta_0(X, U)\|^2 + (1 - \omega) \|\delta_{\lambda_q}^{soft}(X, U) - \theta\|^2 \right)
$$
\[
\begin{align*}
&= \frac{1}{\sigma^2} E \left[ (\omega^3 - \omega + 1) \left( U^T U \right)^2 \|g(X)\|^2 + \|\gamma_q(X)\|^2 \\
&\quad + 2U^T U(\omega^2 - \omega + 1)\gamma_q^T(X)g(X) \\
&\quad + (1 - \omega) \left( \|X - \theta\|^2 + 2\nabla \gamma_q(X) + \frac{2}{k + 2} \left( U^T U \right)^2 \nabla g(X) \right) \right].
\end{align*}
\]

\[
\begin{align*}
R^2 &= \frac{1}{\sigma^2} E \left( \omega \|\delta_{\lambda,q}^\text{soft}(X, U) - \delta_0^\text{soft}(X)\|^2 + (1 - \omega) \|\delta_{\lambda,q}^\text{soft}(X, U) - \theta\|^2 \right) \\
&= \frac{1}{\sigma^2} E \left( (U^T U)^2 \|g(X)\|^2 + 2U^T U(\omega^2 - \omega + 1)\gamma_q^T(X)g(X) \\
&\quad + (\omega^3 - \omega + 1)\|\gamma_q(X)\|^2 \\
&\quad + (1 - \omega) \left( \|X - \theta\|^2 + 2\nabla \gamma_q(X) + \frac{2}{k + 2} \left( U^T U \right)^2 \nabla g(X) \right) \right].
\end{align*}
\]

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

**Remark 1.** Suppose that \((X, U) \sim SS_p((\theta, 0), \sigma^2 I_p)\) with known \(\sigma^2\). For the shrinkage soft threshold estimator \(\delta_{\lambda,q}^\text{soft}(X, U)\), under the balanced loss \(\text{BEL}(\delta_0)\) in (1.2), the \(\lambda^\text{sure}\) threshold is given by

\[
\lambda^\text{sure} = \arg \min_{0 \leq \lambda \leq \lambda^U} SURE_{(X, \lambda)}(\delta_{\lambda,q}^\text{soft}(X, U), \delta_0),
\]

where the \(SURE_{(X, \lambda)}(\delta_{\lambda,q}^\text{soft}(X, U), \delta_0)\) have the following cases:

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

\[
\frac{(\omega^3 - \omega + 1)}{\sigma^2} \left( U^T U \right)^2 \sum_{i=1}^{p} (|X_i| \wedge \lambda)^2 \\
+ \frac{1}{\sigma^2} \sum_{i=1}^{p} X_i^2 I(X_i < 0) + p(1 - \omega)
\]
\begin{align*}
&+ \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}^{p} I(X_i < 0) - \frac{2(1 - \omega)}{\sigma^2} (U^T U)^2 \sum_{i=1}^{p} I(|X_i| \leq \lambda). \\
\end{align*}

2. For the target estimator \( \delta^*_0(X) \), the risk estimate is equal to
\begin{align*}
&\left(\frac{U^T U}{\sigma^2}\right)^2 \sum_{i=1}^{p} (|X_i| \wedge \lambda)^2 + p(1 - \omega) \\
&+ \frac{2(\omega^2 - \omega + 1)U^T U}{\sigma^2} \sum_{i=1}^{p} \gamma_q(X_i) g(X_i) \\
&+ \frac{2(\omega^3 - \omega + 1)}{\sigma^2} \sum_{i=1}^{p} X_i^2 I(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) (U^T U)^2 \sum_{i=1}^{p} I(|X_i| \leq \lambda).
\end{align*}

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
(2.15) with known \( \Sigma \). We wish to estimate \( \theta = (\theta_1, \ldots, \theta_p)^T \) by
\( \delta^\text{soft,}_q = (\delta_1, \ldots, \delta_p)^T \) under the balanced loss function and again we consider the
cases where a subset of \( \theta_i \geq 0, i = 1, \ldots, p \) are non-negative. The shrinkage
soft threshold and target estimators are respectively defined as follows.
\begin{align*}
\delta^\text{soft,}_q(X) &= X + \gamma_q(X) + g(X), \\
\delta_0(X) &= X + (1 - \omega) g(X).
\end{align*}

Again, our purpose is to find a risk for the shrinkage soft threshold estimator
\( \delta^\text{soft,}_q(X) \) under \( \text{BEL}(\delta_0) \) 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 \( S^{-1} = A = (a_{ij})_{1 \leq i, j \leq p} \).

By using the target estimators \( \delta_0(X) \) and \( \delta'_0(X) \) and by using 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.

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

In the following remark, similar 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 (2.15) with known \(\Sigma\) and \(S = VV^T\). Assume that \(g(X)\) is a function into \(\mathbb{R}^p\) which is weakly differentiable in \(X\) and differentiable in \(S\). For the shrinkage soft threshold estimator \(\delta^{soft}_{\lambda, q}(X)\), under the balanced loss \(BEL(\delta_0)\) in (1.1), the \(\lambda^{sure}\) threshold is given by

\[
\lambda^{sure} = \arg \min_{0 \leq \lambda \leq \lambda^U} SURE_{(X, \lambda)}(\delta^{soft}_{\lambda, q}(X), \delta_0),
\]

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

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

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

2. For the target estimator \( \delta^*_0(X) \) the risk estimate is equal to

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

4. **Simulation study**

In this section, we compared theoretical outcomes with numerical computations and simulations to investigate the performance of the soft wavelet shrinkage estimator. All calculations in this section are done using the R software. For denoising or shrinkage coefficients, one of the most important concepts in wavelets and denoising is using thresholds. Shrinkage of the empirical wavelet coefficients works best in problems where the underlying set of the true coefficients of \( f \) is sparse. The shrinkage wavelet method algorithm is as follows:
1. First, the discrete wavelet transform is derived from the noisy observations. In other word, let $Y_1, \ldots, Y_n$ are observed data from model,

$$Y_i = f(X_i) + \eta_i, \quad (4.21)$$

where the $\{\eta_i\}$ is some noise and $\{X_i\}$ is some points from domain of $f(\cdot)$. Typically $n$ is an integer power of 2. If $W$ represent the discrete wavelet transform matrix, then multiplication of the equation (4.21) by orthogonal matrix $W$ yields:

$$X = WY = Wf + W\eta = \theta + \epsilon.$$ 

Note that the observations are sampled from distribution $f$ but with some noise and we are interest to remove noises. 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 groups of high effect (important) and low effect coefficients. If the wavelet coefficient is greater than the threshold value, it belongs to the category of important coefficients, otherwise the set of coefficients will be negligible. Then the low effect coefficients are removed and the important coefficients with respect to the shrink function are given by follows: In hard shrink function, the coefficients are less than
the threshold value equal to zero and the other coefficients remain unchanged. In soft elimination function, coefficients less than threshold value equal to zero and other coefficients decrease as much as 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 2-variate t-distribution $t_2(\mu, \Sigma)$ where $\mu = (0, 0)^T = 0$ and $\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$. Figure 1 shows the density function of $t_2(\mu, \Sigma)$ and density function of $t_2(\mu, \Sigma)$ with added identical independent distribution (iid) $N(0, 1)$ noise with $N = 128$ (noisy $t_2(0, \Sigma)$) along with their contour plot. Now, by de-noising of the noisy $t_2(0, \Sigma)$, we are looking for the main $t_2(0, \Sigma)$. The minimization over the SURE is simply a grid search. The SURE value is first calculated for all $X$ values, and then we find the $\lambda_{\text{sure}}$ among all these values. Using the SURE threshold, the noisy $t_2(0, \Sigma)$ is converted to wavelet coefficients and denoised coefficients by the inverse discrete wavelet transformation, giving an approximation of the main $t_2(0, \Sigma)$. In Figures 2, 3 and 4, for $N = 128, 256, 512$, the density function and contour plot of the noisy $t_2(0, \Sigma)$ signal after the SURE threshold de-noising are plotted.
Figure 1: Top left: Density function of a $t_2(0, \Sigma)$ with $N = 128$. Top right: Density function of a $t_2(0, \Sigma)$ with added iid $N(0, 1)$ noise with $N = 128$. Bottom left: Contour plot of a $t_2(0, \Sigma)$ with $N = 128$. Bottom right: Contour plot of a $t_2(0, \Sigma)$ with added iid $N(0, 1)$ noise with $N = 128$.

We compare between the SURE method and four commonly used shrinkage strategies: hard and soft thresholding with the universal threshold, cross validation (CV) and Bayes thresholding. To assess the performance, we calculated the average mean squared error (AMSE) from $n = 1000$ replications.
Figure 2: Density function and contour plot of a noisy $t_2(0, \Sigma)$ signal after new threshold de-noising with $N = 128$.

Figure 3: Density function and contour plot of a noisy $t_2(0, \Sigma)$ signal after new threshold de-noising with $N = 256$.

of simulation. The value of AMSE is obtained as follows:

$$\frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{N} \frac{f(x_i) - \hat{f}(x_{i,j})}{N},$$
Figure 4: Density function and contour plot of a noisy $t_2(0, \Sigma)$ signal after new threshold de-noising with $N = 512$.

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 higher accuracy of the estimate.

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

As can be seen in the tables, by increasing the $\sigma^2$, the accuracy of all methods decrease. Also, the AMSE amount obtained in the SURE method is lower than that of the other methods, in particular by decreasing the amount of dim $p$. In general, with the increase of the value of $\omega$, the
estimation accuracy decreases.

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

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$p$</th>
<th>SURE</th>
<th>Universal $\omega = 0.3$</th>
<th>Universal $\omega = 0.5$</th>
<th>Universal $\omega = 0.8$</th>
<th>Soft</th>
<th>Hard</th>
<th>threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.7303</td>
<td>0.7307</td>
<td>0.7326</td>
<td>0.7141</td>
<td>0.8542</td>
<td>–</td>
<td>0.7617</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.5762</td>
<td>0.5781</td>
<td>0.5894</td>
<td>0.5805</td>
<td>0.7571</td>
<td>0.5797</td>
<td>0.6733</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.4948</td>
<td>0.5006</td>
<td>0.5241</td>
<td>0.5318</td>
<td>0.6581</td>
<td>0.5247</td>
<td>0.5997</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.4511</td>
<td>0.4607</td>
<td>0.4870</td>
<td>0.5165</td>
<td>0.5761</td>
<td>0.5068</td>
<td>0.5464</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.4293</td>
<td>0.4416</td>
<td>0.4715</td>
<td>0.5073</td>
<td>0.5410</td>
<td>0.4998</td>
<td>0.5213</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.8756</td>
<td>0.8742</td>
<td>0.8707</td>
<td>0.8811</td>
<td>0.9642</td>
<td>–</td>
<td>0.9426</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.7860</td>
<td>0.7765</td>
<td>0.7630</td>
<td>0.8682</td>
<td>0.9745</td>
<td>0.8422</td>
<td>0.9352</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.7394</td>
<td>0.7233</td>
<td>0.7038</td>
<td>0.9229</td>
<td>0.9938</td>
<td>0.8952</td>
<td>0.9584</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>0.6968</td>
<td>0.6749</td>
<td>0.6514</td>
<td>0.9660</td>
<td>0.9879</td>
<td>0.9409</td>
<td>0.9772</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>0.6668</td>
<td>0.6443</td>
<td>0.6247</td>
<td>0.9830</td>
<td>0.9954</td>
<td>0.9657</td>
<td>0.9885</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.2249</td>
<td>1.2170</td>
<td>1.1991</td>
<td>1.2172</td>
<td>1.1801</td>
<td>–</td>
<td>1.3089</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.2523</td>
<td>1.2212</td>
<td>1.1877</td>
<td>1.4512</td>
<td>1.4216</td>
<td>1.3699</td>
<td>1.4551</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1.2364</td>
<td>1.2026</td>
<td>1.1440</td>
<td>1.7153</td>
<td>1.6719</td>
<td>1.6387</td>
<td>1.6986</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1.1730</td>
<td>1.1330</td>
<td>1.0618</td>
<td>1.8651</td>
<td>1.8133</td>
<td>1.8104</td>
<td>1.8420</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>1.1241</td>
<td>1.0791</td>
<td>1.0131</td>
<td>1.9345</td>
<td>1.9032</td>
<td>1.8917</td>
<td>1.9210</td>
<td></td>
</tr>
</tbody>
</table>

We calculated the AMSE for target estimators in non-negative location
parameter space from 1000 replications of simulation. Table 2 represent the AMSE with respect to $p$, $\sigma^2$ and $\omega$ for target estimators for the $N_2(0, \sigma^2 I_2)$. In this table, we can see the following.

- The AMSE value of $\delta_0^{(1)}(X)$ for the corresponding values is lower than that of $\delta_0^{(2)}(X)$.
- By increasing the $\sigma^2$, the accuracy of all methods decrease.
- By increasing the value of dim $p$, the precision of the estimator is increased.
- By increasing of the value of $\omega$, the estimation accuracy decreases.

Figures 5 and 6 show the risk curve of the shrinkage soft threshold estimator (3.20) for non-negative location parameter space for multivariate normal distribution $N_p(0, 2I_p)$ with $p = 16$. These figures were plotted with respect to target estimators $\delta_0(X, U)$ and $\delta_0^*(X)$ for $q = 6, 10$ and $\omega = 0.2, 0.5, 0.8$. Also, Figures 5 and 6 show that the risk is decreased by increasing the amount of $\omega$. By increasing the value of $q$ from 6 to 10, the difference in the risk of the estimators decreases with each other. In general, the risk values in the target estimator $\delta_0^*(X)$ are lower than the target estimator $\delta_0(X, U)$.
Table 2: AMSE for target estimators.

<table>
<thead>
<tr>
<th>$\sigma^2$</th>
<th>$p$</th>
<th>$\omega = 0.3$</th>
<th>$\omega = 0.5$</th>
<th>$\omega = 0.8$</th>
<th>$\delta_0^{(1)}(X)$</th>
<th>$\delta_0^{(2)}(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.7407843</td>
<td>0.7519258</td>
<td>0.8225013</td>
<td>$0.7526715$</td>
<td>0.7759163</td>
<td>0.8344153</td>
</tr>
<tr>
<td>8</td>
<td>0.5957935</td>
<td>0.6122736</td>
<td>0.7792235</td>
<td>$0.606251$</td>
<td>0.6403176</td>
<td>0.8134959</td>
</tr>
<tr>
<td>0.5</td>
<td>0.4993819</td>
<td>0.5126939</td>
<td>0.5041723</td>
<td>$0.5041723$</td>
<td>0.5375324</td>
<td>0.8085041</td>
</tr>
<tr>
<td>32</td>
<td>0.4561108</td>
<td>0.4626690</td>
<td>0.7590374</td>
<td>$0.4577497$</td>
<td>0.4710982</td>
<td>0.8202242</td>
</tr>
<tr>
<td>64</td>
<td>0.4452455</td>
<td>0.4394154</td>
<td>0.7608320</td>
<td>$0.4484006$</td>
<td>0.4470308</td>
<td>0.8313430</td>
</tr>
<tr>
<td>4</td>
<td>0.8114424</td>
<td>0.7984361</td>
<td>0.8405045</td>
<td>$0.8118768$</td>
<td>0.8103646</td>
<td>0.8509157</td>
</tr>
<tr>
<td>8</td>
<td>0.7162636</td>
<td>0.7110568</td>
<td>0.8133852</td>
<td>$0.7180052$</td>
<td>0.7234461</td>
<td>0.8324369</td>
</tr>
<tr>
<td>1</td>
<td>0.661435</td>
<td>0.6648597</td>
<td>0.8010723</td>
<td>$0.6699519$</td>
<td>0.6887543</td>
<td>0.8348478</td>
</tr>
<tr>
<td>32</td>
<td>0.6443929</td>
<td>0.6328510</td>
<td>0.7926348</td>
<td>$0.6554189$</td>
<td>0.6594202</td>
<td>0.8386649</td>
</tr>
<tr>
<td>64</td>
<td>0.6331861</td>
<td>0.6178847</td>
<td>0.7813184</td>
<td>$0.6505679$</td>
<td>0.6450380</td>
<td>0.8312735</td>
</tr>
<tr>
<td>4</td>
<td>0.9136572</td>
<td>0.8933877</td>
<td>0.8961540</td>
<td>$0.9043011$</td>
<td>0.8834847</td>
<td>0.8984375</td>
</tr>
<tr>
<td>8</td>
<td>0.8717649</td>
<td>0.8417345</td>
<td>0.8602557</td>
<td>$0.8579672$</td>
<td>0.8360454</td>
<td>0.8681166</td>
</tr>
<tr>
<td>2</td>
<td>0.8301697</td>
<td>0.7927979</td>
<td>0.8357576</td>
<td>$0.8183334$</td>
<td>0.7947373</td>
<td>0.8558379</td>
</tr>
<tr>
<td>32</td>
<td>0.8228262</td>
<td>0.7835409</td>
<td>0.8329445</td>
<td>$0.8042471$</td>
<td>0.7827455</td>
<td>0.8580542</td>
</tr>
<tr>
<td>64</td>
<td>0.8224749</td>
<td>0.7801745</td>
<td>0.8206946</td>
<td>$0.8023012$</td>
<td>0.7719847</td>
<td>0.8491778</td>
</tr>
</tbody>
</table>

5. Conclusion

In this paper, we generalized the SURE threshold for elliptical and spherical multivariate distributions under the balanced loss function. Also we
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 $\omega$.

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 $\omega$. 
A New Wavelet Threshold SURE of Elliptically Distribution

found the SURE threshold for non-negative mean vector of these distributions. The performance of the shrinkage soft threshold estimator with SURE thresholding was investigated using a simulation study. The results of the figures and tables illustrate the target estimator is appropriate and by increasing the sample size, the accuracy of the new shrinkage estimator increases. Also, the average mean square error of the shrinkage soft threshold estimator is lower than others.

References


REFERENCES

Multivariate Analysis, 9, 531-544.


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