Stat Comput (2009) 19: 27–34
DOI 10.1007/s11222-008-9066-y

Robust wavelet shrinkage using robust selection of thresholds

Hee-Seok Oh · Donghoh Kim · Yongdai Kim

Received: 28 June 2007 / Accepted: 8 April 2008 / Published online: 8 May 2008

Abstract This paper considers the problem of selecting a robust threshold of wavelet shrinkage. Previous approaches reported in literature to handle the presence of outliers mainly focus on developing a robust procedure for a given threshold; this is related to solving a nontrivial optimization problem. The drawback of this approach is that the selection of a robust threshold, which is crucial for the resulting fit is ignored. This paper points out that the best fit can be achieved by a robust wavelet shrinkage with a robust threshold. We propose data-driven selection methods for a robust threshold. These approaches are based on a coupling of classical wavelet thresholding rules with pseudo data. The concept of pseudo data has influenced the implementation of the proposed methods, and provides a fast and efficient algorithm. Results from a simulation study and a real example demonstrate the promising empirical properties of the proposed approaches.

Keywords Cross-validation · Pseudo data · Robust smoothing · SURE · Wavelets

H.-S. Oh (✉) · Y. Kim
Department of Statistics, Seoul National University, Seoul 151-747, Korea
e-mail: heeseok@stats.snu.ac.kr

Y. Kim
e-mail: donghohkim@sejong.ac.kr

D. Kim
Department of Applied Mathematics, Sejong University, Seoul 143-747, Korea
e-mail: ydkim@stats.snu.ac.kr

1 Introduction

Suppose that we observe a set of noisy data satisfying

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n = 2^J, \tag{1}$$

where the ϵ_i's are independently and identically distributed (i.i.d.) random errors and f is an unknown function of interest. The distribution of the errors is typically assumed to be Gaussian.

1.1 An example and main contribution

However, when noise has a non-normal distribution, for instance, heavy-tailed distribution, the classical wavelet shrinkage may not be efficient for estimating the true function as shown in Fig. 1. Interestingly, this figure reflects a motivation of this study that *real* robust wavelet denoising requires a robust procedure as well as a robust threshold. From Fig. 1, it is clear that neither the fit by a classical wavelet denoising with a robust threshold nor the fit by a robust procedure with a non-robust threshold captures the feature of the data. The result from the former is too sensitive to outliers, whereas the fit by the latter is over-smoothed. Indeed, the best fit is obtained from a robust wavelet procedure with a robust threshold. The robust wavelet procedure and the robust criterion of threshold used in Fig. 1 will be described in Sects. 2 and 3.

In literature, developing a robust procedure for wavelet denoising has been mainly focused, but the selection of a robust threshold seems to have been neglected. The objective of this study is to overcome this problem when outliers are present by proposing a robust wavelet denoising procedure with the *automatic robust selection* of the threshold. The main contribution of the paper is to introduce data-driven robust selection methods for the wavelet threshold. We suggest

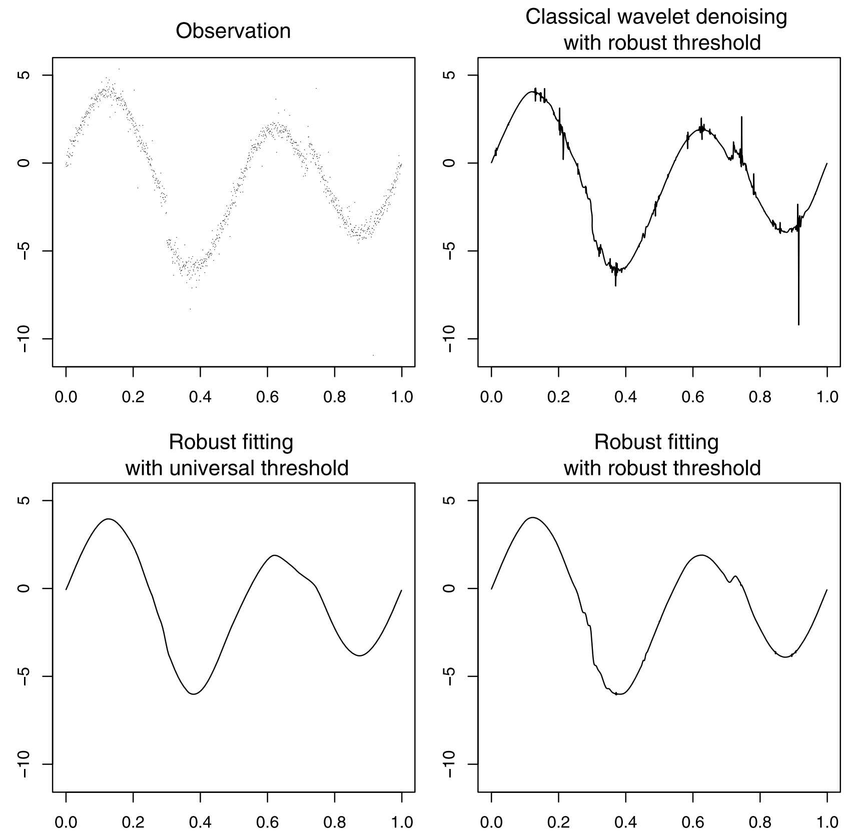

Fig. 1 Estimation when outliers are present

two robust selection techniques by the coupling of a robust version of cross-validation (Nason 1996), and SureShrink (Donoho and Johnstone 1995) with the concept of *pseudo data*. They are related to a robust prediction error criterion that considers the predictive performance for the data, which will be discussed in Sect. 3.

1.2 Previous work

Some robust wavelet procedures have been previously studied. Bruce et al. (1994) proposed a robust wavelet transform to estimate wavelet coefficients at each resolution level which is robust to outliers. Kovac and Silverman (2000) developed a direct approach. This approach can be summarized as identifying outliers with a classical robustness test, removing the outliers, and applying their thresholding procedure for unequally spaced data. However, this approach has a drawback in that information is lost by removing data. Sardy et al. (2001) proposed a robust wavelet denoising estimator based on a robust M-type loss function. More recently, Oh et al. (2007) studied a robust wavelet shrinkage algorithm that is based on the concept of pseudo data. Both procedures are identical for the L_1 penalty and computationally efficient. However, in both procedures, the selection of a robust threshold was not suitably addressed.

As a study of the threshold selection for various types of noise, Averkamp and Houdré (2003) extended the minimax theory of wavelet thresholding to some broader class of *known* symmetric heavy tail noises. This method, therefore, is not always applicable in practice as the distribution of noise is usually unknown.

In this paper, we use the method of Oh et al. (2007) as a robust wavelet denoising procedure. It will be used not only to fit the data for a given threshold value but also to implement robust threshold criteria. The reason for such a selection is that this procedure is fast, simple to implement, and easy to extend to high-dimensional cases. However, the proposed selection methods for a robust threshold can be coupled with any other robust wavelet denoising methods.

2 Review: robust wavelet shrinkage

Let $\boldsymbol{f}$ be the underlying function collected at all dyadic points $\{x_1, \ldots, x_n\}$. Let $\boldsymbol{W}$ be a discrete and orthogonal wavelet transform matrix and $\boldsymbol{\theta}$ be the vector of the wavelet coefficients of $\boldsymbol{f}$. Since $\boldsymbol{W}$ is an orthogonal matrix, the true signal can be expressed as $\boldsymbol{f} = \boldsymbol{W}^T\boldsymbol{\theta}$. Given a sampled data vector $\boldsymbol{y}$, the model (1) can be expressed in matrix notation as follows:

$$\boldsymbol{y} = \boldsymbol{W}^T\boldsymbol{\theta} + \boldsymbol{\epsilon}.$$

In order to get a good estimator of $\boldsymbol{\theta}$, we consider the penalized least squares problem to determine the minimizer of

$$\|\boldsymbol{y} - \boldsymbol{W}^T\boldsymbol{\theta}\|^2 + \lambda \sum_{i=1}^{n} p(|\theta_i|) \tag{2}$$

for a given penalty function and a threshold λ. A typical L_1-penalty, $p(|\theta|) = |\theta|$ leads to the soft-thresholding rule (Donoho et al. 1992). For the selection of λ, various approaches have been studied. For details, see Vidakovic (1999), and Percival and Walden (2000).

However, it is well known that the minimizer of (2) is sensitive to the presence of outliers. One way to overcome this problem is to replace the sum of squares in (2) by a robust loss function. In other words, the robust estimate is the minimizer of

$$\sum_{i=1}^{n} \rho(y_i - \boldsymbol{f}_i) + \lambda \sum_{i=1}^{n} p(|\theta_i|),$$

where $\boldsymbol{f}_i = (\boldsymbol{W}^T\boldsymbol{\theta})_i$. The function $\rho(t)$ is chosen to downweight outlying observations. Typically, $\rho(t)$ is symmetric about zero, quadratic in the neighborhood of zero, and increasing at a rate slower than t^2 for large t. A favorite choice of $\rho(t)$ is the Huber loss function defined as

$$\rho(t) = \begin{cases} t^2, & \text{if } |t| \le c; \\ c(2|t| - c), & \text{otherwise,} \end{cases}$$

where c is a cutoff point that is usually determined from the data. Throughout this study, the cutoff point has been fixed at $c = 1.345$MAD, which ensures 95% efficiency with respect to the normal model in a location problem.

Now, we define pseudo data introduced by Cox (1983) as

$$\tilde{y}_i = f(x_i) + \frac{\psi\{y_i - f(x_i)\}}{2},$$

where $\psi = \rho'$ is the (almost everywhere) derivative of ρ. The concept of pseudo data is the main part of the proposed selection methods. However, the pseudo data $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^T$ are not available as it involves the unknown function. For a practical algorithm, Oh et al. (2007) introduced a fixed point analogy to pseudo data, which is the *empirical pseudo data*

$$z_i = \hat{f}(x_i) + \frac{\psi\{y_i - \hat{f}(x_i)\}}{2},$$

where $\hat{f}$ denotes an estimate of f. In this paper, the robust wavelet estimate $\hat{f}$ is obtained by the ES algorithm of Oh et al. (2007) based on empirical pseudo data.

However, as shown in Fig. 1, the choice of λ is very crucial for the quality of a robust wavelet estimate. Therefore, we propose two data-driven robust selection methods for the threshold.

Before closing this section, we emphasize that the proposed methods described in the next section are not limited to the above-mentioned robust wavelet shrinkage procedure. They can be implemented with any other robust wavelet estimator such as those by Bruce et al. (1994) and Sardy et al. (2001).

3 Robust selection of threshold

This section presents the proposed methods for the robust selection of the wavelet threshold. These are developed by extending the two-fold cross-validation (TFCV) of Nason (1996), and SureShrink (SURE) of Donoho and Johnstone (1995).

3.1 Robust selection based on cross-validation

Cross-validation is one of the most popular choice for selecting the smoothing parameter λ in nonparametric regression (Green and Silverman 1994), it provides a direct method of determining the thresholds without relying on the noise assumption. Cross-validation is defined as

$$\mathrm{CV}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{f}_\lambda^{-i}(x_i)\}^2,$$

where $\hat{f}_\lambda^{-i}(x_i)$ denotes the fit at x_i without the ith observation of the data.

However, when the distribution of errors in (1) is heavy-tailed, the expectation of $\mathrm{CV}(\lambda)$ may not be finite. Then any smoothing parameter chosen from $\mathrm{CV}(\lambda)$ does not have any meaning. In addition, it is computationally expensive to evaluate the fitted values of the robust estimate $\hat{f}_\lambda^{-i}(x_i)$ at each step.

To overcome the above problems, we consider a natural robust extension of the CV as follows

$$\mathrm{RCV}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\rho\{y_i - \hat{f}_\lambda^{-i}(x_i)\},$$

where the robust estimates are obtained from a fast robust estimator such as ES-algorithm. Assume that $\hat{\lambda}$ is an optimal choice obtained by $\mathrm{RCV}(\lambda)$. The $\hat{\lambda}$ will satisfy $-\sum_{i=1}^{n}\psi\{y_i - \hat{f}_\lambda^{-i}(x_i)\}|_{\lambda=\hat{\lambda}} = 0$. Then by the definition of empirical pseudo data, it follows that

$$\begin{aligned}\psi\{y_i - \hat{f}_{\hat{\lambda}}^{-i}(x_i)\} &\approx \psi\{y_i - \hat{f}_{\hat{\lambda}}(x_i)\} \\ &= 2\{z_i - \hat{f}_{\hat{\lambda}}(x_i)\} \approx 2\{z_i - \hat{f}_{\hat{\lambda}}^{-i}(x_i)\}.\end{aligned}$$

Thus the $\hat{\lambda}$ is in fact the minimizer of the following quantity

$$\mathrm{PCV}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\{z_i - \hat{f}_\lambda^{-i}(x_i)\}^2.$$

Note that the only difference between CV(λ) and PCV(λ) is that the original data y_i is replaced by the corresponding empirical pseudo data z_i.

In the context of wavelet shrinkage, PCV(λ) cannot be directly computed due to the limitation of Mallat's fast algorithm for a discrete wavelet transform, which requires the length of the data to be dyadic. Fortunately, Nason (1996) has already tackled this problem based on CV (λ). This method is termed as two-fold cross-validation (TFCV). We now propose a robust method for selecting the threshold λ by simply applying TFCV to PCV(λ).

Let us consider the data $y_1, \ldots, y_n$ where $n = 2^J$. We assume that some outliers are present in the data. With odd-indexed data y_i, we obtain a wavelet estimate $\hat{f}_\lambda^{odd}(x_j)$ by a particular threshold λ from the re-indexed y_j ($j = 1, \ldots, n/2$) and then calculate the corresponding empirical pseudo data $z_{\lambda,j}^{odd}$. To compare the wavelet estimator with the removed data, we form the following interpolation:

$$\tilde{f}_\lambda^{even}(x_j) = \frac{1}{2}\left\{\hat{f}_\lambda^{odd}(x_{j+1}) + \hat{f}_\lambda^{odd}(x_j)\right\}$$

for $j = 1, 2, \ldots, n/2$. For even-indexed data, we compute $\hat{f}_\lambda^{even}(x_j)$, $z_{\lambda,j}^{even}$ and $\tilde{f}_\lambda^{odd}(x_j)$ for a fixed λ in the same manner as above. Then, we construct an approximate PCV(λ) with interpolated estimators and empirical pseudo data

$$\begin{aligned}\mathrm{PTFCV}(\lambda) = \frac{1}{n}\sum_{j=1}^{n/2}\big[&\{z_{\lambda,j}^{odd} - \tilde{f}_\lambda^{odd}(x_j)\}^2 \\ &+ \{z_{\lambda,j}^{even} - \tilde{f}_\lambda^{even}(x_j)\}^2\big].\end{aligned}$$

Let λ^* be the minimizer of PTFCV(λ). However, the λ^* value is based on $n/2$ data points, thus a correction is necessary. By just following the suggestion of Nason (1996), the final threshold is

$$\hat{\lambda} = \left(1 - \frac{\log 2}{\log n}\right)^{-1/2} \lambda^*.$$

In this paper, the implementation of PTFCV is mainly performed by the ES algorithm. That is, with odd-indexed data, the ES algorithm simultaneously gives the converged empirical pseudo data $z_{\lambda,j}^{odd}$ and the corresponding wavelet estimate $\hat{f}_\lambda^{odd}(x_j)$ for a given λ. Then, we easily obtain the interpolation fit $\tilde{f}_\lambda^{even}(x_j)$. Similarly, $z_{\lambda,j}^{even}$ and $\tilde{f}_\lambda^{odd}(x_j)$ can be obtained.

3.2 Robust selection based on SURE

In this section, we propose a robust selection of threshold based on SureShrink (SURE) and empirical pseudo data. SURE is a technique of selecting a threshold by minimizing Stein's unbiased estimator of risk. Let $\tilde{\boldsymbol{f}}_\lambda = \boldsymbol{W}^T\tilde{\boldsymbol{\theta}}_\lambda$ be a wavelet shrinkage estimator, where $\tilde{\boldsymbol{\theta}}_\lambda = \delta_\lambda(\boldsymbol{d})$ represents the thresholding procedure with a threshold λ, and $\boldsymbol{d} = \boldsymbol{W}\boldsymbol{y}$ denotes the empirical wavelet coefficients. The quantity

$$\begin{aligned}\mathrm{SURE}(\boldsymbol{d}(\boldsymbol{y}), \lambda) = n - 2\sum_{i=1}^{n} I(|d_i| \le \lambda) \\ + \sum_{i=1}^{n}(|d_i| \wedge \lambda)^2 \qquad (3)\end{aligned}$$

is an unbiased estimate of $\mathrm{Risk}(\tilde{\boldsymbol{\theta}}_\lambda) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\tilde{\theta}_{\lambda,i} - \theta_i)^2$. Here, $a \wedge b = \min(a, b)$. Due to the orthogonality of $\boldsymbol{W}$, SURE is also an unbiased estimate of $\mathrm{Risk}(\tilde{\boldsymbol{f}}_\lambda)$. Then we find a threshold to minimize SURE over a range of λ as

$$\tilde{\lambda} = \operatorname*{argmin}_{0 \le \lambda \le \sqrt{2\log n}} \mathrm{SURE}(\boldsymbol{d}(\boldsymbol{y}), \lambda).$$

We finally note that it is easy to show that $\mathrm{Risk}(\tilde{\boldsymbol{f}}_\lambda) = \mathrm{PSE}(\tilde{\boldsymbol{f}}_\lambda) + \sigma^2$, where

$$\mathrm{PSE}(\tilde{\boldsymbol{f}}_\lambda) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\{y_i^* - \tilde{f}_\lambda(x_i)\}^2. \qquad (4)$$

Thus, the above SURE procedure is closely related to the minimization of $\mathrm{PSE}(\tilde{\boldsymbol{f}}_\lambda)$.

We now consider an extension of SURE for a robust choice of the threshold. Let $\hat{\boldsymbol{f}}_\lambda$ be a robust wavelet estimator with threshold λ. The robust risk is defined as $\mathrm{Risk}(\hat{\boldsymbol{f}}_\lambda) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\{\hat{f}_\lambda(x_i) - f(x_i)\}^2$, and the corresponding $\mathrm{PSE}(\hat{\boldsymbol{f}}_\lambda)$ can be expressed as $\mathrm{PSE}(\hat{\boldsymbol{f}}_\lambda) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\{\tilde{y}_i^* - \hat{f}_\lambda(x_i)\}^2$, where $\tilde{y}_i^*$ denotes an independent copy of the pseudo data $\tilde{y}_i = f(x_i) + \psi(\epsilon_i)/2$ that can be considered as a transformation of robust additive model with bounded errors. Then it follows that

$$\mathrm{Risk}(\hat{\boldsymbol{f}}_\lambda) = \mathrm{PSE}(\hat{\boldsymbol{f}}_\lambda) + \sigma_{pseudo}^2,$$

where $\sigma_{pseudo}^2 = \mathrm{Var}\{\psi(\epsilon)/2\} < \infty$. Therefore, the relationship between (3) and (4) provides a counterpart of $\mathrm{PSE}(\hat{\boldsymbol{f}}_\lambda)$, that is, a robust version of SURE based on pseudo data as $\mathrm{PSURE}(\lambda) = \mathrm{SURE}(\boldsymbol{d}(\tilde{\boldsymbol{y}}), \lambda)$, where $\boldsymbol{d}(\tilde{\boldsymbol{y}}) = \boldsymbol{W}\tilde{\boldsymbol{y}}$. However, the quantity $\tilde{\boldsymbol{y}}$ is not available practically. Thus, we consider a plug-in-type estimator of PSURE(λ) by simply replacing $\tilde{\boldsymbol{y}}$ with a sample analogue of the pseudo data, $\boldsymbol{z}$ as

$$\begin{aligned}\mathrm{PSURE}(\lambda, h) &= \mathrm{SURE}(\boldsymbol{d}(z_h), \lambda) \\ &= n - 2\sum_{i=1}^{n} I(|d_i(z_{h,i})| \le \lambda) \\ &\quad + \sum_{i=1}^{n}(|d_i(z_{h,i})| \wedge \lambda)^2,\end{aligned}$$

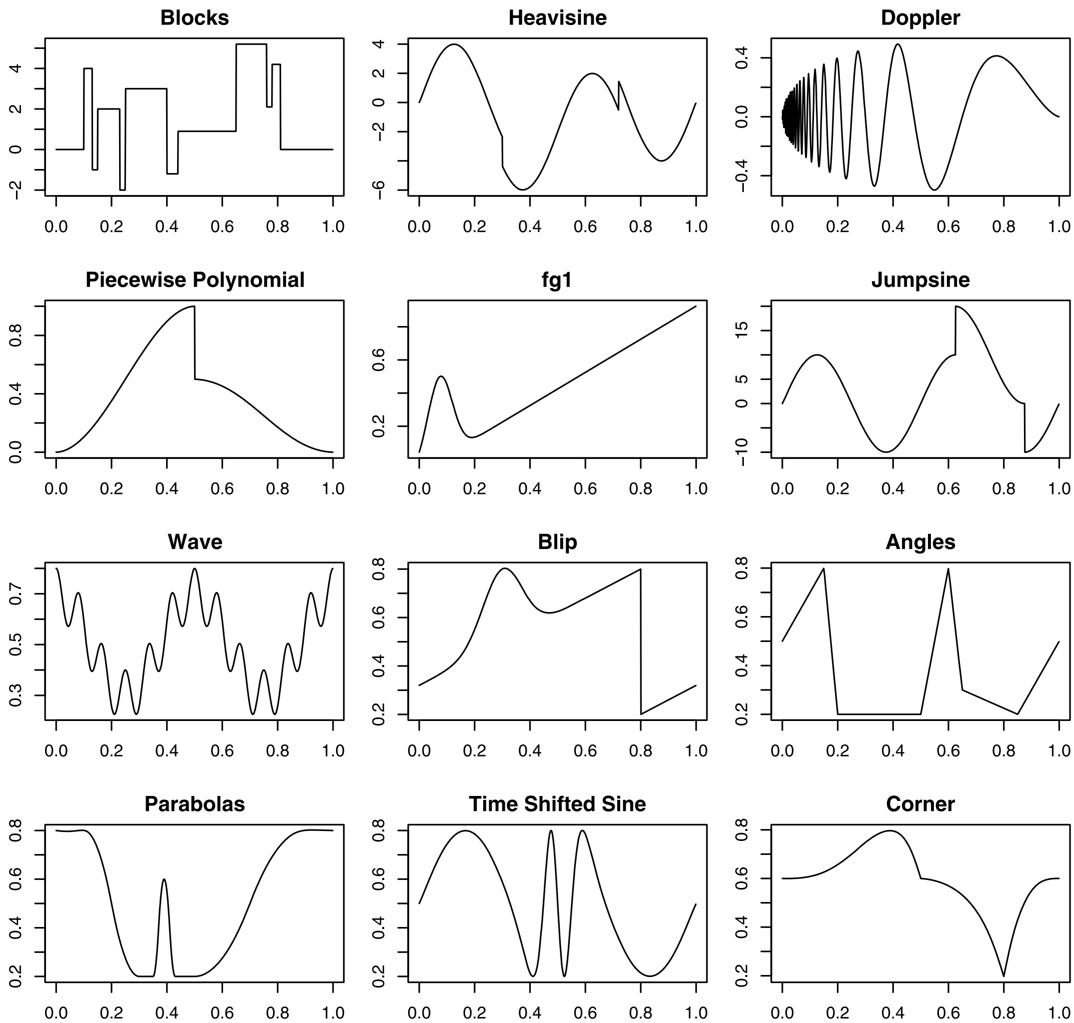

Fig. 2 Test functions

where $\boldsymbol{d}(z_h) = \boldsymbol{W} z_h$ and $z_{h,i} = \hat{f}_h(x_i) + \psi\{y_i - \hat{f}_h(x_i)\}/2$. Note that the threshold h used to define $z_{h,i}$ is not necessary the same as the threshold λ. Since the proposed PSURE depends on two thresholds λ and h, we suggest the following steps to minimize PSURE score:

1. For each threshold h, $\text{PSURE}(h) = \min_\lambda \text{PSURE}(\lambda, h)$ is computed.
2. The threshold $\hat{h}$ is determined by minimizing PSURE(h) for all h.

We state the following remarks with regard to the PSURE procedure. Firstly, for the minimization of PSURE(λ, h), various minimizations can be considered. The main reason to consider the above-mentioned steps is that given a data set $\boldsymbol{d}(z_h)$ from a particular h, SURE$(\boldsymbol{d}(z_h), \lambda)$ is an unbiased estimate of the risk with respect to the data. Moreover, this approach makes the computational burden very light. Secondly, SURE was originally developed for level-dependent thresholding, but, in this paper, a *global* PSURE with one level is applied to the experiments even though it can be extended to a level-wise version.

4 Simulation study and real example

This section reports the results from a simulation study and a real example that are designed to assess the practical performance of the two proposed methods. The codes of the `R` statistical package used to implement the methods and to carry out all the experiments are available at http://stat.snu.ac.kr/heeseok/rt in order that one can reproduce the same results.

4.1 Simulation study

In this study, we compare three non-robust methods and two proposed robust methods for the threshold:

1. `tcv`: the two-fold cross-validation of Nason (1996),

Table 1 Wilcoxon test rankings of the five methods

Test	Normal					CN(0.1)					T(3)				
functions	eby	tcv	mmax	ptfcv	psure	eby	tcv	mmax	ptfcv	psure	eby	tcv	mmax	ptfcv	psure
1	5	1	3.5	2	3.5	5	2	4	2	2	5	1	4	2	3
2	5	1.5	3.5	1.5	3.5	5	4	3	1	2	5	3.5	3.5	1	2
3	5	1	4	2.5	2.5	5	3	4	1.5	1.5	5	3	4	1.5	1.5
4	5	2.5	2.5	2.5	2.5	5	4	3	1	2	5	4	3	1.5	1.5
5	5	2.5	2.5	2.5	2.5	5	3.5	3.5	1.5	1.5	5	2.5	4.0	1	2.5
6	5	2.5	2.5	2.5	2.5	5	4	3	1	2	5	3.5	3.5	1.5	1.5
7	5	1.5	4	1.5	3	5	2	4	2	2	5	2	4	2	2
8	5	2.5	2.5	2.5	2.5	5	3.5	3.5	1	2	5	3	4	1	2
9	5	2	4	2	2	5	3	4	1.5	1.5	5	3	4	1	2
10	5	2.5	2.5	2.5	2.5	5	4	3	1.5	1.5	5	3.5	3.5	1.5	1.5
11	5	2	4	2	2	5	3.5	3.5	1.5	1.5	5	3.5	3.5	1.5	1.5
12	5	2.5	2.5	2.5	2.5	5	4	3	1.5	1.5	5	3	4	1.5	1.5
Mean	5.00	2.00	3.17	2.21	2.63	5.00	3.38	3.46	1.42	1.75	5.00	2.96	3.75	1.42	1.88

Test functions 1: Blocks, 2: Heavisine, 3: Doppler, 4: Piecewise Polynomial, 5: fg1, 6: Jumpsine, 7: Wave, 8: Blip, 9: Angles, 10: Parabolas, 11: Time Shifted Sine, and 12: Corner

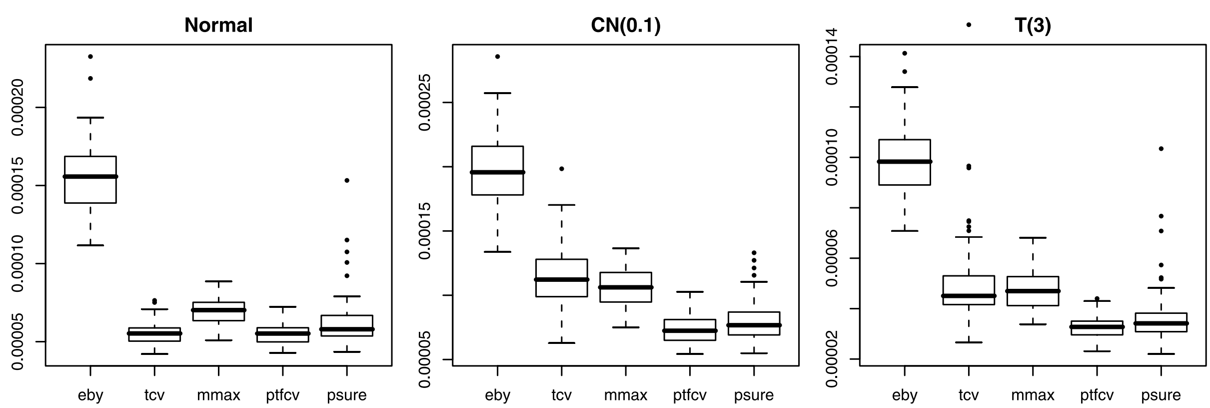

Fig. 3 Boxplots of MSE($\hat{f}_\lambda$) with Time Shifted Sine test function and three noise types

2. mmax: the minimax selection (Donoho and Johnstone 1994) and $c = 2.0$MAD, which is the suggestion of Sardy et al. (2001),
3. eby: the threshold selected by empirical Bayes procedure of Johnstone and Silverman (2005),
4. ptfcv: the proposed method described in Sect. 3.1, and
5. psure: the proposed method in Sect. 3.2.

The following three types of i.i.d. noise of ϵ's are considered:

- Normal: a normal noise from $N(0, \sigma^2_{snr=7})$,
- CN(0.10): a 10% contaminated normal mixture distribution $0.9N(0, \sigma^2_{snr=7})+0.1N(0, \sigma^2_{snr=7/4})$, and
- T(3): t-distribution with three degrees of freedom.

The last two types represent the case in which outliers are present. We consider twelve test functions as plotted in Fig. 2: Blocks, Heavisine, Doppler from Donoho and Johnstone (1994), Piecewise Polynomial from Nason and Silverman (1994), fg1 from Fan and Gijbels (1995), Jumpsine from Bruce and Gao (1996), Wave, Blip, Angles, Parabolas, Time Shifted Sine from Marron et al. (1998) and Corner from Cai (1999). The sample size is fixed as $n = 1024$.

For each combination of the test function and noise type, altogether 100 sets of samples are generated. For each simulated data set, the above five thresholding methods are applied for selecting the threshold.

For assessing the performance of the method, we calculate the mean-squared error (MSE) defined as $\mathrm{MSE}(\hat{f}_\lambda) = \frac{1}{n}\sum_{i=1}^{n}\{\hat{f}_\lambda(x_i) - f(x_i)\}^2$, where λ is determined by the

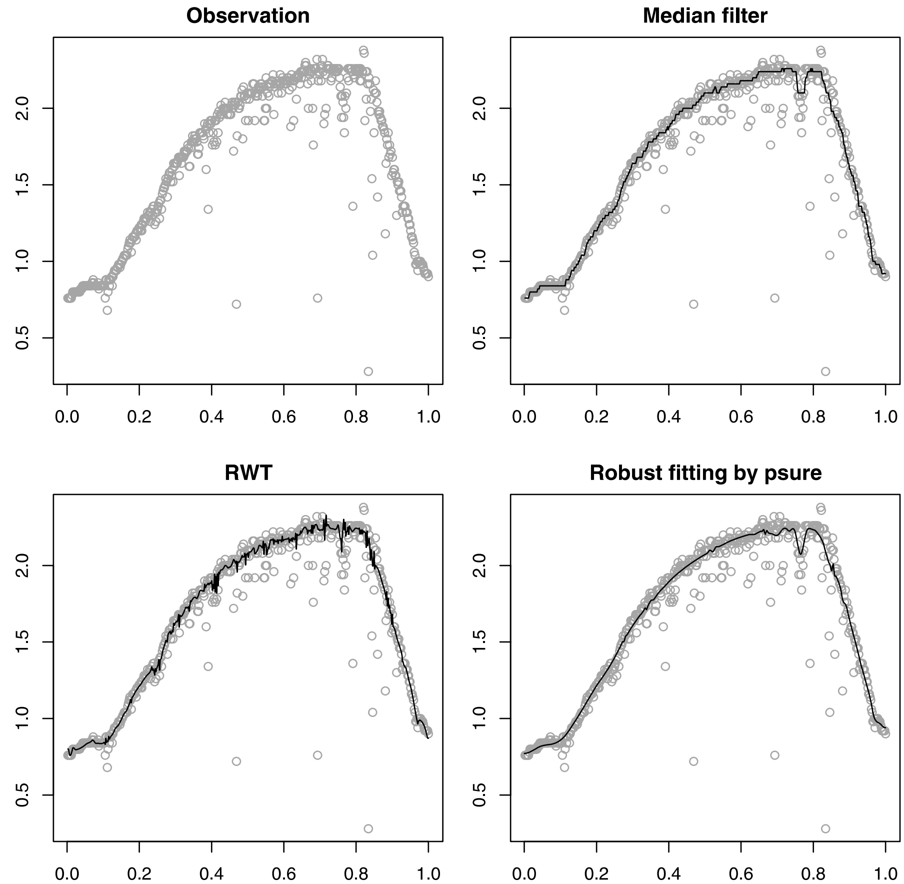

Fig. 4 Robust fitting for balloon data

above five approaches, and the robust fit $\hat{f}_{\hat{\lambda}}$ is obtained by the ES algorithm given $\hat{\lambda}$. Then we follow the procedure used in Lee (2004) for pairwise comparison of the methods. Wilcoxon rank test is applied to test whether the difference between the median MSE values of any two methods is significant or not. The methods are ranked in ascending order with respect to the significant median MSE values. If the median MSE values of the methods are insignificant, the same averaged rank is assigned. The resulting rankings are given in Table 1. Figure 3 displays the boxplots of MSEs for Time Shifted Sine test function with three noise types. Boxplots of other functions are similar, and hence are omitted.

From the simulation results, we obtain the following empirical observations:

1. the results of the robust methods are comparable to those of non-robust ones with regard to normal noise,
2. the robust approaches outperform non-robust approaches for non-Gaussian cases, and
3. for all cases, `ptfcv` gives identical results with `psure`, or `ptfcv` is superior to `psure`.

Overall, the simulation results seem to suggest that the proposed methods are very preferable for the selection of a threshold for wavelet shrinkage when outliers are present.

4.2 Real example

We apply the proposed methods to the balloon data used by Kovac and Silverman (2000). The data are radiation measurements from the sun, performed from a flight of a weather balloon. Since the measurement device was occasionally cut off from the sun, individual outliers and large outlier patches are introduced. We select every 9th or 10th observation from the 4,984 observations to reduce the sample size to 512. Although the data are very closely located to regular grid, but these are not exactly equally spaced. Thus, we modify the ES algorithm by employing the irregularly spaced wavelet thresholding scheme proposed by Kovac and Silverman (2000) in the smoothing step. Figure 4 shows the balloon data with a sample size of 512 and the estimates by a moving median filter with size 17, the robust wavelet transform (RWT) of Bruce et al. (1994) followed by thresholding with SureShrink, and the robust fitting based on the proposed `psure`, respectively. Note that the window size of the median filter is selected by eye to be “best”. Increasing the size does not help capturing the valley around 0.77 and decreasing the size makes the fit wiggly. The fits by the

RWT are still affected by the outliers. On the other hand, the robust fitting based on `psure` only reacts to the major pattern of the observations. Note that the results by `ptfcv` are similar with those of `psure`, and hence, are omitted.

5 Conclusion

Many existing robust wavelet shrinkage procedures have been proposed for a given threshold. We have pointed out that a robust selection of the threshold is crucial for the quality of fits. In this paper, we have proposed automatic robust methods for selecting the threshold of wavelet shrinkage. The new methods are influenced by the concept of pseudo data, and are implemented by a coupling of classical thresholding methods and conventional wavelet estimates with empirical pseudo data. The results from numerical experiments and a real example suggest that the methods possess promising empirical properties and give comparable or superior mean squared error performance to non-robust methods for various noise types.

Acknowledgements This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2005-003-C00032).

References

Averkamp, R., Houdré, C.: Wavelets thresholding for non necessarily Gaussian noise: Idealism. Ann. Stat. **31**, 110–151 (2003)

Bruce, A., Gao, H.-Y.: Applied Wavelet Analysis with S-PLUS. Springer, New York (1996)

Bruce, A., Donoho, D., Gao, H., Martin, D.: Denoising and robust nonlinear wavelet analysis. SPIE Proc. Wavelet Appl. **2242**, 325–336 (1994)

Cai, T.T.: Adaptive wavelet estimation: a block thresholding and oracle inequality approach. Ann. Stat. **27**, 898–924 (1999)

Cox, D.D.: Asymptotics for M-type smoothing splines. Ann. Stat. **11**, 530–551 (1983)

Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**, 425–455 (1994)

Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothing via wavelet shrinkage. J. Am. Stat. Assoc. **90**, 1200–1224 (1995)

Donoho, D.L., Johnstone, I.M., Hock, J.C., Stern, A.S.: Maximum entropy and the nearly black object. J. R. Stat. Soc. Ser. B **54**, 41–81 (1992)

Fan, J., Gijbels, I.: Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. J. R. Stat. Soc. Ser. B **57**, 371–394 (1995)

Green, P.J., Silverman, B.W.: Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman and Hall, London (1994)

Johnstone, I.M., Silverman, B.W.: Empirical Bayes selection of wavelet thresholds. Ann. Stat. **33**, 1700–1752 (2005)

Kovac, A., Silverman, B.W.: Extending the scope of wavelet regression methods by coefficient-dependent thresholding. J. Am. Stat. Assoc. **95**, 172–183 (2000)

Lee, T.C.M.: Improved smoothing spline regression by combining estimates of different smoothness. Stat. Probab. Lett. **67**, 133–140 (2004)

Marron, J.S., Adak, S., Johnstone, I.M., Neumann, M.H., Patil, P.: Exact risk analysis of wavelet regression. J. Comput. Graph. Stat. **7**, 278–309 (1998)

Nason, G.P.: Wavelet shrinkage by cross-validation. J. R. Stat. Soc. Ser. B **58**, 463–479 (1996)

Nason, G.P., Silverman, B.W.: The discrete wavelet transform in S. J. Comput. Graph. Stat. **3**, 163–191 (1994)

Oh, H.-S., Nychka, D.W., Lee, T.C.M.: The role of pseudo data for robust smoothing with application to wavelet regression. Biometrika **94**, 893–904 (2007)

Percival, D.B., Walden, A.T.: Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge (2000)

Sardy, S., Tseng, P., Bruce, A.: Robust wavelet denoising. IEEE Trans. Signal Process. **49**, 1146–1152 (2001)

Vidakovic, B.: Statistical Modeling by Wavelets. Wiley, New York (1999)