Data Adaptive Median Filters for Signal and Image Denoising Using a Generalized SURE Criterion

Hsin-Cheng Huang and Thomas C. M. Lee, *Senior Member, IEEE*

***Abstract*—Due to its effectiveness for removing heavy-tail noise and preserving abrupt structures hidden in noisy data, median filtering has long been a popular tool for signal restoration. In practice, an important issue of applying median filtering is the choice of the span. In this letter, we develop a data adaptive criterion for choosing this span. This criterion is derived using the generalized SURE technique recently proposed by Shen and Huang. It is designed to handle outliers and heavy-tail noise, and it aims to minimize the mean-squared error between the true and restored signals. Results from simulation experiments suggest that the proposed criterion is superior to its competitors.**

***Index Terms*—Covariance penalty, cross-validation, Gaussian mixtures, generalized Stein's unbiased risk estimation (GSURE), outlier modeling, robust smoothing.**

I. Introduction

MEDIAN filtering is a nonlinear robust technique for recovering signals hidden in noisy data. For the following two situations, median filtering tends to provide better signal estimates when compared to linear estimators. The first is when the signals to be recovered possess discontinuities. A typical example for this is image denoising, as most real images contain many sharp edges. The second situation is when the noise distribution is heavy-tail or when the data are contaminated by outliers. Consequently, median filtering has been a subject of active research (see, e.g., [2]–[8]).

In the practical application of median filtering, an important ingredient is the choice of the span of the moving window, i.e., to choose a suitable amount of smoothing. It is because, if the span is too small, spurious features will not be removed from the resulting signal estimates, while if the span is too big, real structures of the signals will be smeared out. While the span can be chosen satisfactorily in a subjective manner, often, objective methods are more desirable. To the best of our knowledge, most existing span selection methods found in the literature are cross-validation (CV) based (see, e.g., [8] and [9]). In this letter, we depart from CV and follow the *covariance penalty* (CP) approach of [1] to develop a span selection method for median filtering. It has been shown that, for other smoothing problems, CP outperforms CV (e.g., [10] and [11]), and hence, the expectation is that CP will also outperform CV for the current span selection problem. Indeed, numerical results reported below support this claim.

The rest of this letter is organized as follows. First, background material is presented in Section II. Then, in Section III, the CP approach is applied to develop methods for span selection. Finally, simulation results are reported in Section IV, while concluding remarks are offered in Section V.

Manuscript received December 15, 2005; revised February 27, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Olivier Besson.

H.-C. Huang is with the Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan, R.O.C. (e-mail: hchuang@stat.sinica.edu.tw).

T. C. M. Lee is with the Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877 USA (e-mail: tlee@stat.colostate.edu).

Digital Object Identifier 10.1109/LSP.2006.874463

II. Background

Consider the following nonparametric regression model: $Y_i = \mu_i + \varepsilon_i$ for $i = 1, \ldots, n$, where $\mu(\cdot)$ is the regression function to be estimated, $Y_1, \ldots, Y_n$ are observations, and $\varepsilon_1, \ldots, \varepsilon_n$ are independently and identically distributed (i.i.d.) random errors with zero mean and variance σ^2.

For all i, denote the median filter estimate with span $2k + 1$ for μ_i as $\hat{\mu}_i^{(k)}$. This estimate is defined as the median of $\{Y_{i-k}, \ldots, Y_i, \ldots, Y_{i+k}\}$. That is, for all i, $\hat{\mu}_i^{(k)} \equiv \mathrm{med}\{Y_{i-k}, \ldots, Y_i, \ldots, Y_{i+k}\}$, where various boundary appending strategies can be applied to define the "data" that are outside the domain of μ (i.e., $Y_{1-k}, \ldots, Y_0$ and $Y_{n+1}, \ldots, Y_{n+k}$). Commonly used strategies include the first and the last values carry-on appending strategy (i.e., $Y_i = Y_1$ for $i \le 1$ and $Y_i = Y_n$ for $i \ge n$) and the periodic appending strategy (i.e., $Y_i = Y_{i-n}$ for $i \in \mathbf{Z}$).

Let $\boldsymbol{\mu} \equiv (\mu_1, \ldots, \mu_n)'$ and $\hat{\boldsymbol{\mu}}^{(k)} \equiv (\hat{\mu}_1^{(k)}, \ldots, \hat{\mu}_n^{(k)})'$. The goal is to select the best value of k so that the following L_2 risk is minimized: $R_k \equiv E\{(\hat{\boldsymbol{\mu}}^{(k)} - \boldsymbol{\mu})'(\hat{\boldsymbol{\mu}}^{(k)} - \boldsymbol{\mu})\}$. In the above, the expectation is taken with respect to the density of $\boldsymbol{Y}$ given $\boldsymbol{\mu}$. As $\boldsymbol{\mu}$ is an unknown quantity, in practice, R_k cannot be minimized directly. One common strategy to solve this problem is to construct an estimator for R_k and choose the k that minimizes such a risk estimator. Stein's unbiased risk estimation [12] (SURE) is a popular technique for constructing such an estimator when the errors ε_i's are i.i.d. Gaussians. However, this SURE technique cannot be directly applied for the case of heavy-tail noise or when outliers are present. To overcome this problem, we shall follow the CP approach and use the generalized SURE (GSURE) formula introduced by [1] to develop such an estimator.

III. Span Selection Using GSURE

Let $\boldsymbol{Y} \equiv (Y_1, \ldots, Y_n)'$. Using the following identity as in [13] $(\hat{\boldsymbol{\mu}}^{(k)} - \boldsymbol{\mu})'(\hat{\boldsymbol{\mu}}^{(k)} - \boldsymbol{\mu}) + (\boldsymbol{Y} - \boldsymbol{\mu})'(\boldsymbol{Y} - \boldsymbol{\mu}) = (\boldsymbol{Y} - \hat{\boldsymbol{\mu}}^{(k)})'(\boldsymbol{Y} - \hat{\boldsymbol{\mu}}^{(k)}) + 2(\hat{\boldsymbol{\mu}}^{(k)} - \boldsymbol{\mu})'(\boldsymbol{Y} - \boldsymbol{\mu})$, one can show that, up to a constant,

R_k can be unbiasedly estimated by $(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}^{(k)})'(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}^{(k)})+2C_k$, where

$$\begin{aligned} C_k &\equiv E\left\{\left(\hat{\boldsymbol{\mu}}^{(k)}-\boldsymbol{\mu}\right)'(\boldsymbol{Y}-\boldsymbol{\mu})\right\} \\ &= E\left[\left\{\hat{\boldsymbol{\mu}}^{(k)}-E\left(\hat{\boldsymbol{\mu}}^{(k)}\right)\right\}'(\boldsymbol{Y}-\boldsymbol{\mu})\right] \\ &= \sum_{i=1}^{n} cov\left(\hat{\mu}_i^{(k)}, Y_i\right) \end{aligned}$$

is the covariance penalty for the median regression estimator $\hat{\boldsymbol{\mu}}^{(k)}$. Of course, in practice, C_k is an unknown, and hence, the above expression cannot be used for estimating R_k.

To apply the GSURE formula for constructing an unbiased estimator for C_k, define

$$V_i(Y_i) \equiv \begin{cases} -\frac{\int_{-\infty}^{Y_i}(y_i-\mu_i)p_i(y_i)dy_i}{p_i(Y_i)}, & \text{if } p_i(Y_i)>0 \\ 0, & \text{otherwise} \end{cases}$$

where $p_i(\cdot)$ is the probability density function of Y_i. Note that $V_i(Y_i)$ is also a function of μ_i, but for simplicity, we suppress this dependence in its notation. Suppose that $p_i(\cdot)$ is almost differentiable in the sense of [12] and that $E|\hat{\mu}_i^{(k)}(Y_i-\mu_i)|<\infty$. Then, it is shown in [1] that $E\{V_i(Y_i)\} = var(Y_i)$ and that $cov(\hat{\mu}_i^{(k)}, Y_i) = E\{V_i(Y_i)(\partial/\partial Y_i)\hat{\mu}_i^{(k)}\}$. Notice that $\hat{\mu}_i^{(k)}$ is continuous and also almost differentiable in Y_i with

$$\frac{\partial}{\partial Y_i}\hat{\mu}_i^{(k)} = \begin{cases} 1, & \text{if } \hat{\mu}_i^{(k)} = Y_i \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Thus $cov(\hat{\mu}_i^{(k)}, Y_i)$ and hence C_k can be estimated via the estimation of $V_i(Y_i)$. Below, we will develop estimators for $V_i(Y_i)$ under two situations: when the i.i.d. errors ε_i's are 1) Gaussians and 2) mixtures of two Gaussian distributions. The latter is a model for data contaminated by outliers.

To sum up, if we denote our final estimate for C_k as $\hat{C}_k$, then our proposed span selection method for median regression is to select the span k that minimizes

$$\left(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}^{(k)}\right)'\left(\boldsymbol{Y}-\hat{\boldsymbol{\mu}}^{(k)}\right)+2\hat{C}_k. \tag{2}$$

A. Gaussian Noise

Suppose that the i.i.d. errors $\{\varepsilon_i; i=1,\ldots,n\}$ are Gaussians, i.e., $Y_i \sim N(\mu_i, \sigma^2)$ for all i. In this case, a direct calculation shows that $V_i(Y_i)=\sigma^2$. Therefore, if an unbiased estimate $\hat{\sigma}^2$ for σ^2 is available, C_k can be unbiasedly estimated by

$$\hat{C}_k \equiv \sum_{i=1}^{n}\hat{\sigma}^2\frac{\partial}{\partial Y_i}\hat{\mu}_i^{(k)} = \hat{\sigma}^2 p \tag{3}$$

where p is the number of Y_i such that $Y_i = \hat{\mu}_i^{(k)}$. This p, which is also the number of observation points Y_i that are passed through by the signal estimates $\hat{\boldsymbol{\mu}}^{(k)}$, can be seen as a measure of model complexity of $\hat{\boldsymbol{\mu}}^{(k)}$. If it is known that the noise $\{\varepsilon_i; i=1,\ldots,n\}$ are i.i.d. Gaussians, we propose to choose the span k as the minimizer of (2) with $\hat{C}_k$ given by (3).

In the 1-D setting with i.i.d. Gaussian noise, various methods have been proposed for estimating σ^2 (see, e.g., [14] and [15]). However, these methods may not be statistically optimal when outliers are present. In Section III-C, we present an effective method for estimating σ^2. This method was designed to accommodate outliers and can be applied to both the 1-D and 2-D settings.

We remark that the above estimator (3) can also be derived using the original SURE formula in [12]. However, for the modeling of outliers or heavy-tail noise, such as the situation to be considered next, SURE cannot be applied.

B. Gaussian Mixtures for Outliers and Heavy-Tail Noise Modeling

Now suppose that the errors $\{\varepsilon_i : i=1,\ldots,n\}$ follow a mixture Gaussian distribution $wN(0,\sigma_1^2)+(1-w)N(0,\sigma_2^2)$, where $\sigma_2^2 > \sigma_1^2$, and $1-w<0.5$ is the percentage of outliers. This is known as the inflated-variance model for outliers [16]. In this case, one can show that

$$V_i(Y_i) = \frac{w\sigma_1\phi\left(\frac{Y_i-\mu_i}{\sigma_1}\right)+(1-w)\sigma_2\phi\left(\frac{Y_i-\mu_i}{\sigma_2}\right)}{w\sigma_1^{-1}\phi\left(\frac{Y_i-\mu_i}{\sigma_1}\right)+(1-w)\sigma_2^{-1}\left(\frac{Y_i-\mu_i}{\sigma_2}\right)}$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution. Thus, if preliminary estimates $\hat{\mu}_i$, $\hat{w}$, $\hat{\sigma}_1$, and $\hat{\sigma}_2$ for, respectively, μ_i, w, σ_1^2, and σ_2^2 are available, a natural plug-in type estimate for $V_i(Y_i)$ is

$$\hat{V}_i(Y_i) \equiv \frac{\hat{w}\hat{\sigma}_1\phi\left(\frac{Y_i-\hat{\mu}_i}{\hat{\sigma}_1}\right)+(1-\hat{w})\hat{\sigma}_2\phi\left(\frac{Y_i-\hat{\mu}_i}{\hat{\sigma}_2}\right)}{\hat{w}\sigma_1^{-1}\phi\left(\frac{Y_i-\hat{\mu}_i}{\hat{\sigma}_1}\right)+(1-\hat{w})\hat{\sigma}_2^{-1}\phi\left(\frac{Y_i-\hat{\mu}_i}{\hat{\sigma}_2}\right)}. \tag{4}$$

In the statistics literature, the above preliminary estimates are sometimes known as *pilot estimates*. Successful uses of these pilot estimates for dealing with various nonparametric estimation problems have been widely reported (see, e.g., [17], [18], and the many references given therein). Methods for obtaining pilot estimates for the current problem are provided in the next subsection.

If it is known that there are outliers or the noise is heavy-tail, we propose to choose k as the minimizer of (2) with $\hat{C}_k$ as

$$\hat{C}_k \equiv \sum_{i=1}^{n}\hat{V}_i(Y_i)\frac{\partial}{\partial Y_i}\hat{\mu}_i^{(k)} \tag{5}$$

where $\hat{V}_i(Y_i)$ and $(\partial/\partial Y_i)\hat{\mu}_i^{(k)}$ are given by (4) and (1), respectively.

Of course, in practice, we will never know if the errors are i.i.d. Gaussians or heavy-tail or if outliers are present. In other words, we will not know if we should estimate C_k by (3) or (5). Our extensive numerical experience suggests that, if the noise is

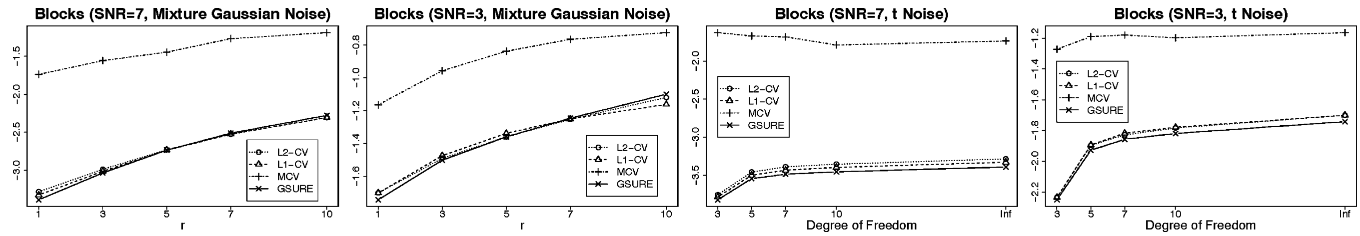

Fig. 1. MSE(k) performance of various methods for reconstructing the *Blocks* function based on 1000 simulation replications.

in fact i.i.d. Gaussians, then estimating C_k with (3) gives marginally better results than using (5). However, if the noise is heavy-tail or outliers are present, the use of (3) produces much worse results then (5). For this reason, we recommend using (5) if no prior knowledge is available about the nature of the noise. Also, in the simulation section below, for the same reason, we only report results when C_k is estimated by (5).

C. Pilot Estimation

This subsection presents our methods for obtaining various pilot estimates. For the pilot estimation of μ_i required in (4), we found that using median filtering with a relatively small k is a reliable and computationally fast choice. Thus, for the 1-D setting, we recommend using $k = 2$ (i.e., a moving window of length 5), and for image data, we recommend $k = 1$ (i.e., a moving window of size 3 × 3). Below, we denote this pilot estimate as $\hat{\mu}_i^*$.

For the estimation of σ_1 as needed in (4), we first compute the pilot error estimate $\hat{e}_i = Y_i - \hat{\mu}_i^*$ for all i. Let $\bar{e}$ and s be, respectively, the average and the median absolute deviation (MAD) of $\hat{e}_1, \ldots, \hat{e}_n$. Remove any $\hat{e}_i$ that is outside the interval $(\bar{e} - 2.58s, \bar{e} + 2.58s)$. This interval can be seen as a robust 99% confidence interval if the $\hat{e}_i$ are i.i.d. Gaussians, as in this case, s is a robust estimate for σ_1, while 2.58 is the 99.5% percentile for the standard Gaussian distribution. Thus, the goal of this step is to remove any $\hat{e}_i$ that is potentially originated from an outlier. Finally, the estimate $\hat{\sigma}_1$ for σ_1 is taken as the MAD of those surviving $\hat{e}_i$'s, i.e., those inside $(\bar{e} - 2.58s, \bar{e} + 2.58s)$. We also use the same procedure to obtain an estimate for σ for the use of (3).

For the estimation of σ_2 in (4), we only use those $\hat{e}_i$'s that are *outside* the interval $(\bar{e} - 2.58s, \bar{e} + 2.58s)$. Let $|\tilde{e}|$ be the average of the absolute values of these "outside $\hat{e}_i$'s". It is straightforward to show that the method-of-moments estimate $\hat{\sigma}_2$ of σ_2 is the solution to the following equation:

$$\log\left(\frac{\sigma_2}{\sqrt{2\pi}}\right) - \frac{(2.58s)^2}{2\sigma_2^2} - \log\left\{1 - \Phi\left(\frac{2.58s}{\sigma_2}\right)\right\} = \log\left(|\tilde{e}|\right)$$

where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution. Notice that the above equation can be solved rapidly using any standard iterative methods.

Last, for w in (4), we use the following method-of-moments estimate:

$$\hat{w} = 1 - \min\left\{\frac{\max\left\{(\hat{e}_i - \bar{e})^2/(n-1) - \hat{\sigma}_1^2, 0\right\}}{\hat{\sigma}_2^2 - \hat{\sigma}_1^2}, 1\right\}.$$

IV. Simulations

Numerical experiments for both 1-D and 2-D settings were conducted to examine the practical performance of the above mixture GSURE criterion for selecting k, i.e., choose the k that minimizes (2) with C_k estimated by (5). Below, we shall refer to this method as GSURE, and it was compared to three CV-based methods: L_2-CV, L_1-CV, and the median CV (MCV) of [9]. Let $\hat{\mu}_{-i}^{(k)}$ be the leave-one-out estimate of μ_i in the 1-D setting, i.e., $\hat{\mu}_{-i}^{(k)} \equiv \mathrm{med}\{Y_{i-k}, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_{i+k}\}$. Then, L_2-CV, L_1-CV, and MCV choose the k that minimizes, respectively,

$$\mathrm{cv}_2(k) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\mu}_{-i}^{(k)}\right)^2, \quad \mathrm{cv}_1(k) = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{\mu}_{-i}^{(k)}\right|$$

and

$$\mathrm{cv}_m(k) = \mathrm{med}\left\{\left|Y_1 - \hat{\mu}_{-1}^{(k)}\right|, \ldots, \left|Y_n - \hat{\mu}_{-n}^{(k)}\right|\right\}.$$

For the 1-D setting, the *blocks* function of [19] was used as the true function μ with $n = 200$. Two types of noise distributions were considered: the Gaussian mixture model discussed in Section III-B and the t-distribution that exhibits heavy-tail behaviors for small degrees of freedom (df). For the mixture model, the data were generated with r = 1, 3, 5, 7, and 10, where $\sigma_2^2 = r^2\sigma_1^2$, and the mixture parameter was fixed at $w = 0.9$. For the t-distribution, the data were generated with $\mathrm{df} = 3$, 5, 7, 10, and ∞. Note that $r = 1$ and $\mathrm{df} = \infty$ correspond to the i.i.d. Gaussian noise case. We considered two signal-to-noise ratios (SNRs): 3 and 7. Let $\bar{\mu} = \sum_{i=1}^{n}\mu_i/n$. For the mixture model, the SNR is defined as $\sqrt{\sum(\mu_i - \bar{\mu})^2/(n\sigma_1^2)}$, while for the t-distribution, it is defined as $\sqrt{\sum(\mu_i - \bar{\mu})^2/(n\sigma_t^2)}$, where σ_t^2 is the variance of the t-distribution under consideration.

For each generated data set, we applied L_2-CV, L_1-CV, MCV, and GSURE to choose their best k and calculated their $\hat{\mu}_i^{(k)}$'s. We have also calculated their corresponding mean-squared error (MSE) values: $\mathrm{MSE}(k) = 1/n\sum_{i=1}^{n}\{\hat{\mu}_i^{(k)} - \mu_i\}^2$. The log of the averages of these MSE(k) values are summarized in Fig. 1.

For the 2-D image denoising experiments, two test images of size 256 × 256 were used. They are the *Lena* image that has been used by many authors and the *Square* image whose pixel value is 1 if the pixel coordinate is inside [65, 192] × [65, 192] and zero otherwise. Other experimental factors (e.g., SNRs, types of noise distributions) were the same as for the 1-D setting. The corresponding log averaged MSE(k) values are summarized in Fig. 2.

From Figs. 1 and 2, the following empirical conclusions can be drawn. First, except for the 1-D mixture noise experiments with $r = 10$, GSURE never gave a worse result than the three

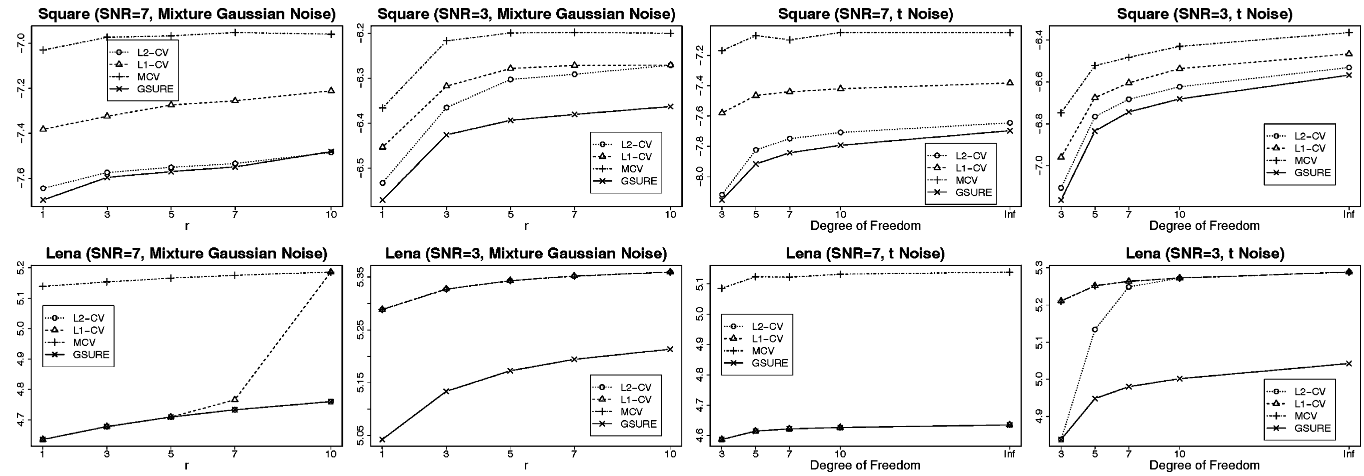

Fig. 2. MSE(k) performance of various methods for reconstructing images based on 100 simulation replications.

CV-based methods. On the other hand, especially for the *Square* image, GSURE often outperformed the CV-based methods. Second, for the t-distribution noise and for all cases with $r = 1$ and $t = \infty$, which correspond to the cases of "pure" Gaussian noise with no outliers nor heavy-tail noise, GSURE performed the best, despite the fact that GSURE was developed under the Gaussian mixture model assumption. Finally, it seems that L_2-CV dominated L_1-CV, while MCV is the least favorable method.

V. Conclusions

In this letter, a new automatic method is developed for choosing the span for median filtering. This new method is designed to handle outliers and heavy-tail noise and is derived using the technique of GSURE. Unlike most GSURE-related procedures for solving nonlinear estimation problems (see, e.g., [1], [10], and [13]), the proposed span selection method does not require such Monte Carlo computations, and hence, it is fast. For example, for the above 1-D experiments, it took less than 1 s to complete 1 replicate on a Pentium-IV 3.2-Ghz PC. Given its speed, this GSURE procedure can also be used as a building block for other more sophisticated median filter methods, such as those cited above. Numerical results from both 1-D and 2-D denoising experiments suggest that this new method is preferable when compared to its competitors.

Last, we conclude this letter with the following general remark about the GSURE methodology. Similar to the original SURE criterion [12] that has been applied to a variety of Gaussian estimation problems, such as estimating the threshold level in wavelet shrinkage (i.e., the SUREShrink of [19]), the GSURE methodology can also be applied to conduct parameter estimation and even to perform model order inference for parametric models. For example, in the time series setting, it has been applied to select a "best" fitting model from a set of different autoregressive models [11].

Acknowledgment

The authors would like to thank the referees for their most constructive comments.

References

[1] X. Shen and H.-C. Huang, "Optimal model assessment, selection and combination," *J. Amer. Statist. Assoc.*, vol. 101, pp. 554–568, 2006.

[2] A. Burian and P. Kuosmanen, "Tuning the smoothness of the recursive median filter," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1631–1639, Jul. 2002.

[3] M. J. Cree, "Observations on adaptive vector filters for noise reduction in color images," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pt. 2, pp. 140–143, Feb. 2004.

[4] H.-L. Eng and K.-K. Ma, "Noise adaptive soft-switching median filter," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 242–251, Feb. 2001.

[5] D. Huang and W. T. M. Dunsmuir, "Computing joint distributions of 2D moving median filters with applications to detection of edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 340–343, Mar. 1998.

[6] L. Khriji and M. Gabbouj, "Vector median-rational hybrid filters for multichannel image processing," *IEEE Signal Process. Lett.*, vol. 6, no. 7, pp. 186–190, Jul. 1999.

[7] G. Qiu, "An improved recursive median filtering scheme for image processing," *IEEE Trans. Image Process.*, vol. 5, no. 4, pp. 646–648, Apr. 1996.

[8] S. J. Reeves, "On the selection of median structure for image filtering," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 42, no. 8, pp. 556–558, Aug. 1995.

[9] Z. G. Zheng and Y. Yang, "Cross-validation and median criterion," *Statistica Sinica*, vol. 8, pp. 907–921, 1998.

[10] X. Shen, H.-C. Huang, and J. Ye, "Adaptive model selection assessment for exponential family models," *Technometrics*, vol. 46, pp. 306–317, 2004.

[11] X. Shen and J. Ye, "Adaptive model selection," *J. Amer. Statist. Assoc.*, vol. 97, pp. 210–221, 2002.

[12] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, pp. 1135–1151, 1981.

[13] B. Efron, "The estimation of prediction error: Covariance penalties and cross-validation (with discussions)," *J. Amer. Statist. Assoc.*, vol. 99, pp. 619–642, 2004.

[14] M. J. Buckley, G. K. Eagleson, and B. W. Silverman, "The estimation of residual variance in nonparametric regression," *Biometrika*, vol. 75, pp. 189–199, 1988.

[15] P. Hall, J. W. Kay, and D. M. Titterington, "Asymptotically optimal difference-based estimation of variance in nonparametric regression," *Biometrika*, vol. 77, pp. 521–528, 1990.

[16] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions*. Chichester, U.K.: Wiley, 1985.

[17] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *J. Amer. Statist. Assoc.*, vol. 91, pp. 401–407, 1996.

[18] T. C. M. Lee, "A stabilized bandwidth selection method for kernel smoothing of the periodogram," *Signal Process.*, vol. 81, pp. 419–430, 2001.

[19] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, 1995.