Journal of Data Science **10**(2012), 259-270

Image De-noising with a New Threshold Value Using Wavelets

B. Ismail[1] and Anjum Khan[2*]
[1]*Mangalore University and* [2] *P. A. College of Engineering*

Abstract: The image de-noising is the process to remove the noise from the image naturally corrupted by the noise. The wavelet method is one among the various methods for recovering infinite dimensional objects like curves, densities, images etc. The wavelet techniques are very effective to remove the noise because of its ability to capture the energy of a signal in few energy transform values. The wavelet methods are based on shrinking the wavelet coefficients in the wavelet domain. This paper concentrates on selecting a threshold for wavelet function estimation. A new threshold value is proposed to shrink the wavelet coefficients obtained by wavelet decomposition of a noisy image by considering that the sub band coefficients have a generalized Gaussian distribution. The proposed threshold value is based on the power of 2 in the size $2^J \times 2^J$ of the data that can be computed efficiently. The experiment has been conducted on various test images to compare with the established threshold parameters. The result shows that the proposed threshold value removes the noise significantly.

Key words: Minimax threshold, orthonormal bases, universal threshold, wavelet shrinkage.

1. Introduction

Wavelets are used widely not only by mathematicians in areas such as functional and numerical analysis but also by researchers in the natural sciences such as physics, chemistry and biology and in applied discipline such as computer science, engineering and econometrics. Statisticians are among the recent users of wavelet applications in areas such as signal processing and image analysis. Signal processing in general, including image analysis and data compression, the use of wavelet has proved of significant value. A comprehensive survey of wavelet application in statistics are given by Ogden (1997). An image is often corrupted by noise during its acquisition or transition. The objective is to remove the noise

*Corresponding author.

without affecting the important feature of the image. The most commonly used procedure to remove the noise is wavelet shrinkage by non-linear method proposed by Donoho and Johnstone (1994, 1995) and Donoho (1995). This approach is now widely used in statistics particularly in signal processing and image analysis. In statistical context this can be referred as the estimation of the true curve from the data contaminated with the noise usually assume to be Gaussian noise. The estimation of the true curve involves three steps. Apply Discrete Wavelet Transform (DWT) which transforms the discrete data from time domain into time-frequency domain. The values of the transformed data in time-frequency domain are called the coefficients. The coefficients with small absolute values dominated by noise, while the coefficients with large absolute values carry more data information than noise. In the second step the wavelet coefficient are set to zero (hard threshold rule) or shrink (soft threshold rule), if they are not crossing certain threshold level. The last step is to reconstruct the signal from the resultant coefficient using Inverse Discrete Wavelet Transform (IDWT). Since the work of Donoho and Johnstone (1994, 1995), there has been much research on finding the threshold value. Nason (1996) obtained a threshold value by cross validation approach, the multiple hypothesis procedure is developed by Abramovich and Benjamin (1995). A small threshold value may yield a result close to the input but the result may still be noisy. A large threshold value on the other hand produce a signal with a large number of zero coefficients, which leads to a smooth signal. Paying too much attention to smoothness however destroys details of the image. In this paper we proposed a threshold value that lies between universal and minimax threshold values (Donoho and Johnstone, 1994) for certain range of data. It approaches to the minimax for data of large size and hence performs well compared to the universal threshold value.

The paper is organized as follows, in Section 2 we provide some necessary mathematical background in relations to wavelets and Discrete Wavelet Transform (DWT). Section 3 introduces the wavelet thresholding procedure. In Section 4 we give a brief review about the choice of threshold and the proposed threshold value. We obtain the experimental results and compare its performance with sure, minimax and universal threshold values through simulations in Section 5. We conclude the paper with some comments in the last section.

2. Some Background on Wavelets

To understand the wavelet applications some of the mathematical background and terminology are presented below:

2.1 Wavelet Series Expansion

The series expansion of a function in terms of the set of orthogonal basis function is familiar in statistics. For example in Fourier expansion the basis consists of sines and cosines of different frequencies. The term wavelets is used to refer to a set of basis functions with a special structure. Many types of functions that are encountered in practice can be sparsely and uniquely represented in terms of the wavelet series. Wavelets are functions specially made as to form an orthonormal basis for various function spaces. One such example is $L^2(R)$, set of all square integrable function on real numbers R. It can be shown (Daubechies, 1992) that it is possible to construct a function $\psi(x)$ so that any function $f \in L^2(R)$ can be represented by

$$f(x) = \sum_{k \in Z} c_{0,k}\phi_{0,k}(x) + \sum_{j<J}\sum_{k \in Z} d_{j,k}\psi_{j,k}(x) \tag{2.1}$$

where $c_{0,k} = \int_R f(x)\phi_{0,k}(x)dx$ and $d_{j,k} = \int_R f(x)\psi_{j,k}(x)dx$, where j controls the maximum resolution. The functions $\psi_{j,k} = 2^{j/2}\psi(2^j x - k)$ are all derived from the mother wavelet $\psi(x)$ by dilation and the translation. The functions $\phi_{0,k}(x)$ are all derived from a function $\phi(x)$ known as father wavelet or scaling function by using dilation and translation formula $\phi_{0,k} = \phi(x-k)$.

The simplest example of wavelet basis is Haar basis (Haar, 1910) which uses scaling function and mother wavelet given by

$$\phi(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{otherwise}, \end{cases}$$

$$\psi(x) = \begin{cases} 1, & 0 \le x < \frac{1}{2}, \\ -1, & \frac{1}{2} \le x \le 1, \\ 0, & \text{otherwise}. \end{cases}$$

In case of two dimension, the scaling function and the wavelets are defined as follows

$$\Phi_{j;k,l}(x,y) = \varphi_{j,k}(x)\varphi_{j,l}(y) = 2^j\Phi(2^j x - k, 2^j y - l),$$
$$\Psi^s_{j;k,l}(x,y) = 2^j\Psi^s(2^j x - k, 2^j y - l),$$

where $s = h, v, d$ are horizontal, vertical and diagonal details respectively defined as,

$$\Psi^h_{j;k,l}(x,y) = \varphi_{j,k}(x)\psi_{j,l}(y),$$
$$\Psi^v_{j;k,l}(x,y) = \psi_{j,k}(x)\varphi_{j,l}(y),$$
$$\Psi^d_{j;k,l}(x,y) = \psi_{j,k}(x)\psi_{j,l}(y).$$

The set $\{\Phi_{j;k,l}(x,y)\} \cup \{\Psi^h_{j;k,l}(x,y), \Psi^v_{j;k,l}(x,y), \Psi^d_{j;k,l}(x,y);\ j;k,l \in Z\}$ is an orthonormal basis for function space $L^2(R^2)$.

Therefore any function $f \in L^2(R^2)$ can be expressed as

$$f(x,y) = \sum_{k,l \in Z} c_{j_0;k,l}\Phi_{j_0;k,l}(x,y) + \sum_i \sum_{j \geq j_0} \sum_{k,l \in Z} d^i_{j;k,l}\Psi^i_{j;k,l}(x,y). \tag{2.2}$$

where $c_{j_0;k,l}$ is scaling coefficient and $d^i_{j;k,l}$ for $i = h, v, d$ are wavelet coefficients, called the sub-band coefficients.

2.2 Discrete Wavelet Transform (DWT)

Usually in statistical problem we have finite set of discrete data. If we have $n = 2^J$ value of $y(x)$ equally spaced between 0 and 1, we use wavelets $\psi_{j,k}(x)$ at levels $j = 0, 1, 2, \cdots, J-1$, where $k = 1, \cdots, 2^j - 1$. Level 0 contains the mother and father wavelets while increasing value of j corresponding to wavelets which describes finer details. If $y = (y(x_1), y(x_2), \cdots, y(x_n))^T$, then

$$y(x_i) = c_{0,0}\phi(x_i) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x_i). \tag{2.3}$$

The vector $w = (c_{0,0}, d_{0,0}, \cdots, d_{J-1,2^j-1})^T$ of coefficients in (2.3) is referred to as the DWT of y. In practice the DWT can be performed using the algorithm of Mallat (1989) with $O(n)$ operation rather than the slow $O(n^2)$ matrix multiplication. In case of two dimensional Gray scale image, which can be thought of as a matrix with entries $x_{i,j}$ corresponding to the intensity of Gray scale in the pixel at position (i,j), the DWT is applied to the rows and column of the matrix. Using Mallats algorithm the process of the DWT goes as follows. The operator H associated with the scaling function called low pass filter and the operator G associated with the wavelet function called high pass filter are applied in the rows of matrix A of size $2^J \times 2^J$. Two resultant matrices H_rA and G_rA are obtained both of size $2^J \times 2^{J-1}$. In the next step the operators are applied on the columns of H_rA and G_rA which results in four matrices H_cH_rA, G_cH_rA, H_cG_rA and G_cG_rA of dimension $2^{J-1} \times 2^{J-1}$, the subscripts r and c suggest that the operations are performed on the rows and columns respectively. The matrix H_cH_rA is an average, while the matrices G_cH_rA, H_cG_rA and G_cG_rA are respectively horizontal, vertical and diagonal details of the matrix. The process continues to obtain the next level decomposition with the matrix H_cH_rA as an input. The Figure 1 shows the schematic representation of two level decomposition.

3. Wavelet Thresholding

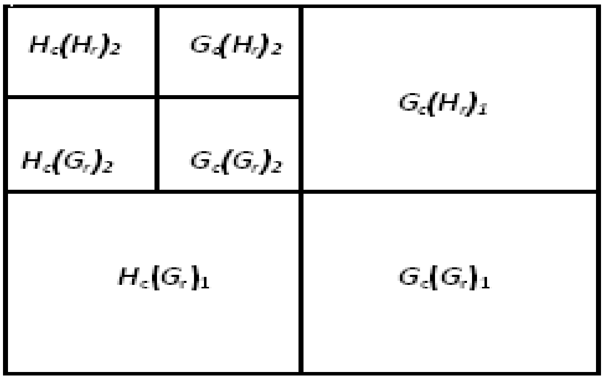

Figure 1: Schematic representation of the visualization of the two dimensional wavelet transform of A

Donoho and Johnstone (1994, 1995) and Donoho *et al.* (1995) proposed a procedure to estimate the curve based on a reconstruction from a more judicious selection of wavelet coefficients. This approach is now widely used in statistics particularly in signal processing and image analysis.

Suppose $\{x_{i,j}, 1 \leq i,j \leq n\}$, is an image of $N = n \times n$ pixels, which is corrupted by white Gaussian noise, then the model for the noisy image is basically of the form:

$$y_{i,j} = x_{i,j} + \epsilon_{i,j}, \qquad 1 \leq i,j \leq n. \tag{3.1}$$

where ϵ is iid as $N(0, \sigma^2)$, σ being the standard deviation of the noise. The goal is to estimate $x_{i,j}$ by removing the noise so that the Mean Square Error is minimum. The estimation process involves three steps as follows,

(1) Apply the two dimensional Discrete Wavelet Transform on the noisy data $y_{i,j}$ to obtain the sub-band approximate coefficients, horizontal details, vertical details and diagonal details. The orthogonal property of the transform insures that the noise in the transform is also of Gaussian nature.

(2) Threshold the detail coefficients using either soft or hard threshold rule. For a given threshold $\lambda > 0$ the hard threshold value is given by

$$\delta^H(w, \lambda) = wI(|w| > \lambda), \tag{3.2}$$

which is a "keep or kill" rule and the soft threshold value is given by

$$\delta^s(w, \lambda) = \text{Sign}(w)(|w| - \lambda)I(|w| > \lambda), \tag{3.3}$$

which is "shrink or kill" rule.

The thresholded wavelet coefficient obtained by applying any of the threshold rule $\delta(w,\lambda)$ given in (3.2) or (3.3) are used to obtain a selective reconstruction.

(3) Reconstruct the image using Inverse Discrete Wavelet Transform of the thresholded wavelet coefficients to obtain the de-noised image $\hat{x}_{i,j}$.

4. The Choice of Threshold

Clearly, an appropriate choice of the threshold value λ is fundamental to the effectiveness of the procedure described in the previous section. Too large threshold might cut off important parts of the true function underlying the data whereas too small a threshold retain noise in a selective reconstruction. Donoho and Johnstone (1995) proposed a sure shrink thresholding rule based on minimizing unbiased risk estimate. The estimate of the risk can be obtained for a particular threshold value λ. Minimising the risk in λ gives a selection of the threshold. The sure shrink is especially designed for soft threshold rule and it is level dependent threshold.

The minimax threshold proposed by Donoho and Johnstone (1994) that depends on the data size N, defined as $\lambda^M = \hat{\sigma}\lambda_N^*$ where λ_N^* is defined as the value of λ which achieves

$$\Lambda_N^* = \inf_{\lambda}\sup_{d}\{R_\lambda(d)/(N^{-1} + R_{oracle}(d))\}, \tag{4.1}$$

where $R_\lambda(d) = E[\delta_\lambda(\hat{d}) - d]^2$ and $R_{oracle}(d)$ is the ideal risk achieved with the help of an oracle.

As an alternative to minimax threshold Donoho and Johnstone (1994) proposed the universal threshold $\lambda_{univ} = \hat{\sigma}\sqrt{2\log N}$, where N is the number of pixels and $\hat{\sigma}$ is estimated standard deviation of the noise. This threshold value is also asymptotically optimal and simpler to implement. It is proved that the maximum of any N values iid as $N(0,\sigma^2)$ will be smaller than universal threshold value with the probability approaching one as N increases. Comparing to the minimax and sure the universal threshold value is substantially large killing many signal coefficients along with the noise.

In this article we have proposed a threshold value which is a fixed form as universal and the minimax threshold. This threshold value depends on the power of two in the size of the data $N = 2^J \times 2^J$ which is defined as $\hat{\sigma}\sqrt{2J}$. In general for n dimensional case this threshold value becomes $\hat{\sigma}\sqrt{nJ}$. The case of $n = 1$ has been used to estimate the true curve from its noisy counterpart (Ismail and Khan, 2010). Since $\hat{\sigma}\sqrt{nJ} \leqslant \lambda_{univ}$ it performs well in terms of the means square error. As the size of the data increase this threshold value gives a better performance. In contrary to the minimax and the universal threshold this proposed threshold works well for the moderate samples, also compare to the sure shrink the performance improves with respect to the increase in the noise level.

In all the above cases, the noise standard deviation can be estimated from the sub-band G_cG_r by the robust median estimator Chang *et al.* (2000) as

$$\hat{\sigma} = \frac{\text{median}(|Y_{i,j}|)}{0.6745}, \quad Y_{i,j} \in \text{sub-band} \;\; G_cG_r.$$

5. Experimental Results and Discussion

The Experiments are conducted on various Gray scale images like Lena, Barbara and Boat of size 512×512 at different noise levels $\sigma = 10, 15, 20, 25$ and 30 using Doubechies (1992) least symmetric compactly supported wavelets with eight vanishing moments at level three. The results are compared with sure, minimax and the universal threshold using both soft and hard threshold rules. The objective quality of the reconstructed image is measured by peak signal to noise ratio

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{\text{MSE}}. \tag{5.1}$$

Where MSE is the mean square error between the original and the de-noised image of size $N = n \times n = 2^J \times 2^J$. The MSE can be calculated using the relation

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n} [x_{i,j} - \hat{x}_{i,j}]^2. \tag{5.2}$$

The Table 1 shows the comparative performance of the sure, universal, minimax and the proposed threshold using the soft threshold rule. The result shows that the proposed threshold value provides better reconstruction for large value of σ. Table 2 shows the comparative performance of the sure, universal, minimax and the proposed threshold using the hard threshold rule. The result shows that at the large values of σ all the values perform almost equally except sure which gives still better performance as it is designed particularly for the soft threshold rule. The proposed threshold value is suitable to both soft and hard threshold rules, where as the sure shrink suitable only to soft threshold.

Figure 2 demonstrate the comparative performance of the sure, universal, minimax and proposed threshold values for Lena image at $\sigma = 30$ with soft threshold rule. Figure 3 demonstrates the comparative performance of sure, minimax,universal and proposed threshold value for Lena image at $\sigma = 30$ with hard threshold rule. Here the performance of sure is worst because sure is designed for soft threshold rule.

Table 1: PSNR values for various noise level σ with soft threshold rule for Lena, Barbara and Boat image respectively

σ	sure	minimax	universal	proposed
10	32.9245	30.0161	28.9805	29.5277
15	30.2887	28.7022	27.8085	28.2762
20	28.2871	27.8163	27.0453	27.4423
25	26.6714	27.1629	26.4963	26.8375
30	25.3169	26.6507	26.0720	26.3641
10	31.1371	26.0320	24.8114	25.4433
15	28.7124	24.6590	23.7043	24.1876
20	26.9065	23.8449	23.1091	23.4678
25	25.4640	23.3155	22.7657	23.0219
30	24.2582	22.9549	22.5467	22.7325
10	31.5441	27.4097	26.3635	26.9113
15	29.2059	26.2163	25.3072	25.7790
20	27.4024	25.4020	24.6121	25.0173
25	25.9308	24.8069	24.1152	24.4660
30	24.6858	24.3427	23.7519	24.0462

Table 2: PSNR values for various noise level σ with hard threshold rule for Lena, Barbara and boat image respectively

σ	sure	minimax	universal	proposed
10	28.8472	31.8928	30.8646	31.4781
15	25.3935	30,2499	29.2870	29,8350
20	22.9281	29.0418	28.2266	28.7236
25	21.0083	28.1535	27.4792	27.8823
30	19.4400	27.4269	26.8143	27.2394
10	28.8160	28.6609	27.0419	27.9284
15	25.3607	26.5521	25.1432	25.9140
20	22.8929	25.2105	24.0332	24.6639
25	20.9737	24.2919	23.3681	23.8262
30	19.4002	23,6157	22,9430	23.2770
10	28.7299	29.4566	28.2434	28.9099
15	25.3104	27.8432	26.8211	27.4033
20	22.8644	26.7455	25.8451	26.3338
25	20.9578	25.9072	25.0933	25.5853
30	19.3958	25.2579	24.5409	24.9230

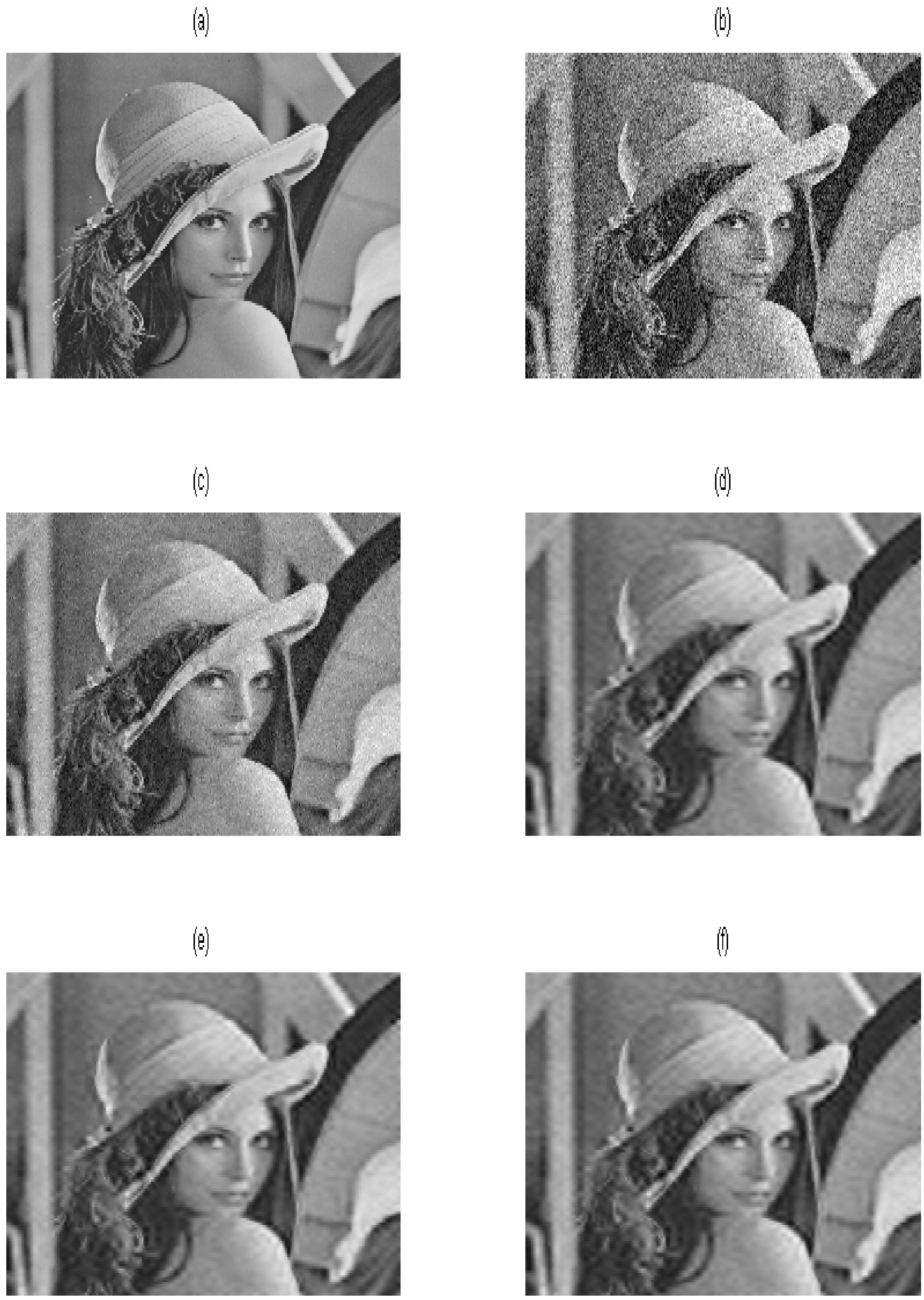

Figure 2: (a) Original Lena image (b) noisy image with $\sigma = 30$ (c) de-noised using sure shrink (d) de-noised using universal (e) de-noised using minimax (f) de-noised using proposed threshold with soft thresholding rule

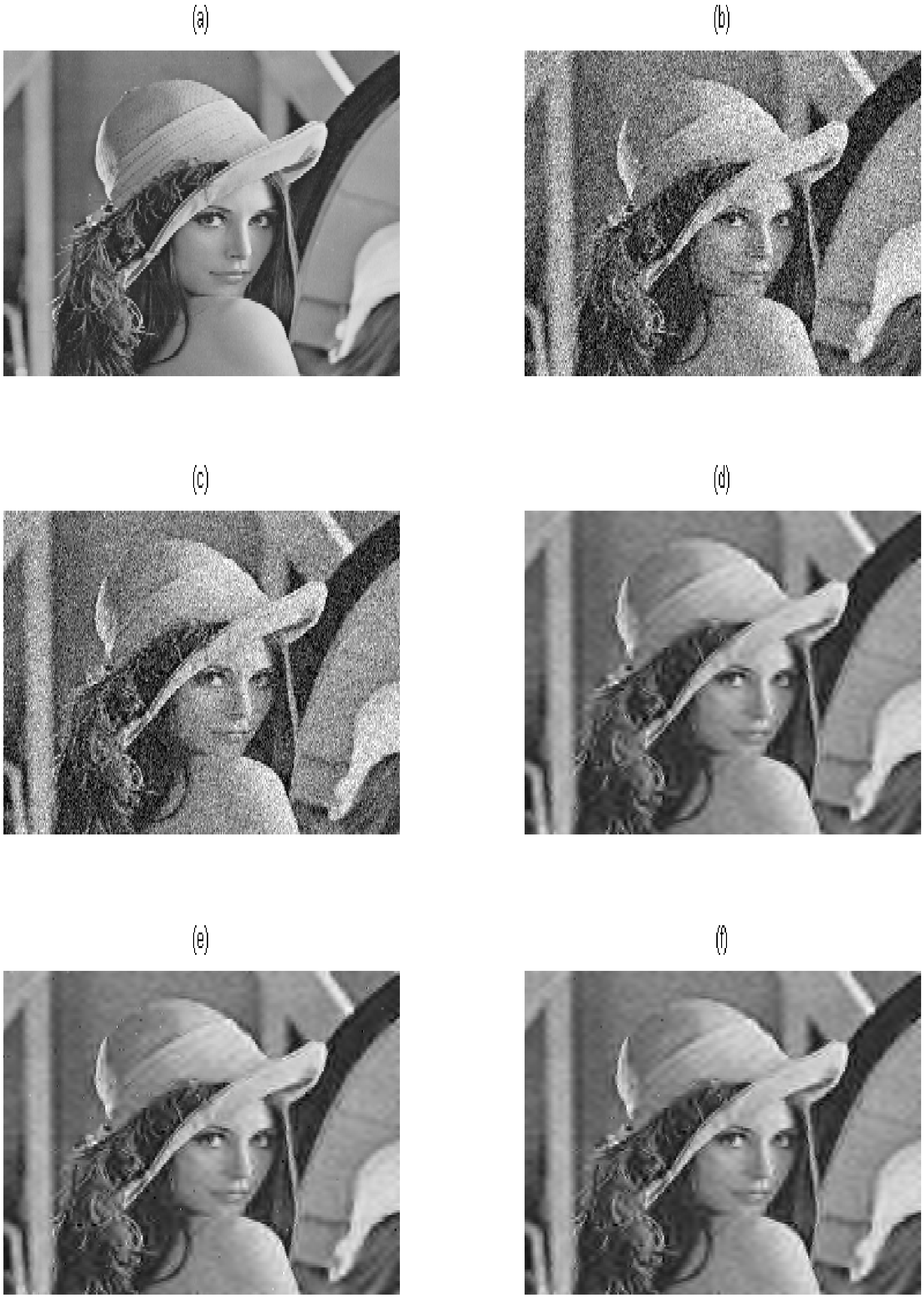

Figure 3: (a) Original Lena image (b) noisy image with $\sigma = 30$ (c) de-noised using sure shrink (d) de-noised using universal (e) de-noised using minimax (f) de-noised using proposed threshold with hard thresholding rule

6. Comments and Conclusion

In this paper a new threshold value is proposed to de-noise the image using wavelets. This threshold value is used to remove the noise from various test images. The result shows that this threshold value removes the noise significantly. We have measured the peak signal to noise ratio for different values of noise level σ to study the performance of the threshold. The comparative result shows that the proposed threshold value found to be better than universal threshold and for large data the proposed threshold value has MSE close to the minimax threshold value. The comparative PSNR value of the proposed threshold improves with increase in the noise level. In this sense our threshold value is an important contribution to the choice of the threshold to remove the noise from the image using wavelets.

Acknowledgements

We are grateful to Prof. Min Te Chao for handling our paper and to the JDS editors. Also we would like to thanks referee for his valuable comments and suggestions which led to an improvement on an earlier version of this paper.

References

Abramovich, F. and Benjamini, Y. (1995). Thresholding the wavelet coefficients as multiple hypothesis testing procedure. In Antoniadis, A. and Oppenheim, G. (Eds.), *Wavelets and Statistics*, Lecture Notes in Statistics **103**, pp. 5-14. Springer-Verlag, New York.

Chang, S. G., Yu, Bin and Vetterli, M. (2000). Adaptive wavelet thresholding for image de-noising and compression. *IEEE Transactions on Image Processing* **9**, 1532-1546.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.

Donoho, D. L. (1995). De-noising by soft thresholding. *IEEE Transaction on Information Theory* **41**, 613-627.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrica* **81**, 425-455.

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

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *Journal of the Royal Statistical Society, Series B* **57**, 301-369.

Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen* **69**, 331-371.

Ismail, B. and Khan, Anjum (2010). A new threshold value in curve estimation by wavelet shrinkage. *Proceedings of the Tenth Islamic Countries Conference on Statistical Sciences* (*ICCS-X*), Volume II, 786-795.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**, 674-693.

Nason, G. P. (1996). Wavelet shrinkage using cross validation. *Journal of the Royal Statistical Society, Series B* **58**, 463-479.

Ogden, R. T. (1997). *Essential Wavelets for Statistical Application and Data Analysis*. Birkhiuser, Boston.

Received December 12, 2010; accepted April 21, 2011.

B. Ismail
Department of Statistics
Mangalore University
Mangalagangothri-574199 India
prof.ismailb@gmail.com

Anjum Khan
P.A. College of Engineering
Kairangala
Near Mangalore University
Mangalore India
khananjum73@gmail.com