Improved Wavelet Denoising by Penalty Function and Regularization Parameter

Wanhui Wei¹,²,³,⁴, Wei Zhou¹,²,³,⁴, * and Yongjun Wu¹,²,³,⁴

¹School of Traffic & Transportation engineering, Central South University, Changsha, Hunan, 410083, China
²Key Laboratory of Traffic Safety on Track, Ministry of Education; Changsha, Hunan, 410083, China
³Joint International Research Laboratory of Key Technology for Rail Traffic Safety, Changsha, Hunan, 410083, China.
⁴National & Local Joint Engineering Research Center of Safety Technology for Rail Vehicle, Changsha, Hunan, 410083, China.

*Corresponding author’s e-mail: 1228945450@qq.com

Abstract. Wavelet-based convex optimization in sparse signal processing has attracted extensive research interest in recent years. Research demonstrates that the penalty function promotes the sparsity and improves the accuracy better than L1 norm method in dealing with convex and sparse signal reconstruction issues. This paper presents an improved denoising model which utilizes a penalty function to induce stronger sparsity in wavelet domain. In this paper, we have redefined the relationship model between signal noise levels and the regularization parameter to enhance sparsity caused by penalty functions in a wavelet domain. And three penalty functions are employed for wavelet coefficients to induce strong sparse wavelet coefficients under the wavelet domain. We then apply the improved wavelet denoising algorithm for image denoising to get the values of the PSNR are analysed under different noise level conditions. Experimental results clearly show that the improved wavelet denoising model is strength in terms of both quantitative measure and structural similarity quality, and outperforms many widely used denoising algorithms.

1. Introduction

Image denoising is a classic problem in signal processing. In the face of practical problems, the noise in the image has a certain impact on the next step of computer image processing. Wavelet analysis can decomposes images into varying wavelets which are component waves of varying durations. These wavelets are localized variations of detail in an image. And they can be used for a wide variety of fundamental image processing tasks, such as removing noise, etc. The two-dimensional wavelet transform can extract the information of the image in the horizontal, vertical and diagonal directions by row transformation and column transformation. This information is very important for expressing the direction information and texture of the image during image processing. Therefore, for images with structural feature laws, the wavelet transform can effectively denoise, which can greatly improve the image performance [1][2][3]. However, the selection of basis function and the determination of threshold function restrict the improvement of performance [4][5].Researchers have made a lot of improvements in the optimization of basis functions, but the basis functions constructed for specific
purposes are narrower and more theoretical. Therefore, most of the wavelet basis functions are now selected based on the L1 norm and entropy in the widely available wavelet basis functions [9] [10][28-31]. The determination of the threshold function is mainly two methods [6] [11][12].

Previous studies have shown that most signals are sparse in the wavelet domain, and wavelet denoising can be regarded as sparse approximation in the wavelet representation domain [6]. In the process of reconstructing signals by sparse approximation, the sparsity of the signal has a great influence on the accuracy of reconstruction. Strong sparsity makes the reconstruction results more accurate. In the wavelet denoising method, the wavelet representation domain generated by a specific wavelet basis function determines the sparsity of wavelet coefficients, and different threshold functions induce different sparsity of wavelet coefficients [7] [8]. In order to improve the sparsity of the reconstructed signal, the researchers designed special penalty functions such as logarithmic function [11], arctangent function [12], minimally large concave function [13] [14], rational function [15] and exponential function [16]. Using these penalty functions, the corresponding threshold function can be derived. These threshold functions have good performance for inducing sparsity [17]. However, there is currently no research on the use of penalty functions to denoise images in the wavelet domain. There is no research to verify that compared with other traditional threshold functions, the wavelet denoising model using the penalty function can obtain better denoising effect as shown in the theory. Moreover, the relationship between the threshold value and the noise variance actually needs further study.

This paper presents an improved wavelet denoising method based on the non-convex penalty function and the regression parameters. The penalty function can induce sparsity more strongly under a wavelet domain. Therefore, we consider employing three penalty functions to improve the performance of the wavelet transform, respectively. Besides, the relationship model between signal noise level and regression parameters is need redefined to strength the sparseness and robustness of the denoising algorithm. The rest of this paper is organized as follows. Section 2 establishes a denoising model based on the different distribution of wavelet coefficient in one wavelet domain. In section 3, we present some experimental tests and the results analysis of different methods. Section 4 concludes this paper.

2. The image denoising algorithm

2.1. Wavelet transform

This section establishes a mathematical model of the noisy image. Models with noisy images are expressed as:

\[ f(x, y) = s(x, y) + n(x, y), 0 < x < M, 0 < y < N \]  

\( f(x,y) \) is the gray value of the point \((x,y)\) of the noisy image of size \(M \times N\), \( s(x,y) \) is the gray value of the point \((x,y)\) in the original image of size \(M \times N\), \( n(x,y) \) is the noise of the point \((x,y)\).

A 2D wavelet transform of a discrete image can be performed by the following steps:

\[ f \rightarrow \begin{pmatrix} h^1 \\ d^1 \\ v^1 \end{pmatrix}, a^1 \rightarrow \begin{pmatrix} h^2 \\ d^2 \\ v^2 \end{pmatrix}, \ldots, a^{n-1} \rightarrow \begin{pmatrix} h^n \\ d^n \\ v^n \end{pmatrix} \]  

A 1-level wavelet transform of an image \( f \) can be divided into four sub images, \( h^1, d^1, a^1 \) and \( v^1 \). Multiple levels of wavelet transform are defined by repeating the 1-level transform of the previous trend, and the fluctuations \( h, d, \) and \( v \) remain unchanged [18]. Image and noise have different statistical characteristics after wavelet transform.
2.2. Image denoising model with penalty function

This section preliminarily established the denoising model under wavelet transform. Wavelet transform can be used to consider signals at different scales as a linear combination of wavelet bases. The observed and original signals can be represented by one wavelet basis function as follows:

\[ f = \omega C_1 \]
\[ s = \omega C_i \]

where \( C_1 \) and \( C_i \) are the wavelet coefficients of the original signal and the observed signal in the wavelet domain, respectively. \( \omega \) represents a wavelet transform matrix based on a basis function.

After the image signal is transformed into the wavelet representation domain, the denoising model based on wavelet transform is defined as:

\[ C = \arg \min_C \left\{ \frac{1}{2} C_j - C^2 + P(C) \right\} \]

Where \( C \) is the number of wavelet coefficients and \( C_j \) is the wavelet coefficient of the observed signal \( f \). The function \( P(C) \) is defined as the penalty function of the wavelet coefficients in the wavelet representation domain, which is given by

\[ P(C) = \sum_{j=1}^J \lambda_j \phi_j(C_j; a) \]

\( \phi_j(C_j; a) \) is a parameter penalty function that promotes sparseness. According to this penalty function, the corresponding threshold function can be obtained. The parameter \( a \) controls the degree to which \( \phi \) is non-convex. Three suitable penalty functions \( \phi \) and their relevant properties as follow [19]:

\[
\phi_j(C_j; a) = \begin{cases} \\
\frac{1}{a} \log\left(1 + a \left| C_j \right|\right), & 0 \leq a \leq \frac{1}{\lambda}, \ \text{log} \\
\frac{2}{a \sqrt{3}} \tan\left(\frac{1 + 2a \left| C_j \right|}{\sqrt{3}}\right), & -\frac{\pi}{6}, a > 0, b > 0, \ \text{tan} \\
\frac{\left| C_j \right|}{1 + a \left| C_j \right|^2}, & a \left| C_j \right| \rightarrow 0, \ \text{rat} \\
\end{cases}
\]

For the logarithmic function, by increasing \( a \) up to \( 1/\lambda \), the gap goes to zero more rapidly. For the logarithmic, arctangent and rational penalty functions, the value of \( a \) can be given by the following formula [16]

\[ a = \beta \frac{\lambda_j}{\lambda_n} \]

According to (7), we then have the new

\[ a = \frac{1}{\lambda_j} \]

The regularization parameter \( \lambda_j \) depends on the exponent \( j \), and it can be obtained by the wavelet scale [20]:

\[ \lambda_j = \frac{\sigma}{2^{j/2}} \]
Where $\sigma$ is the level of noise, $\lambda$ is a constant that usually suggested being 2.5, 3.0, 4.0, etc.

2.3. Improved regularization parameter model

However, for the wavelet denoising method based on a penalty function, the relationship between coefficients is not the same as the soft threshold method. Actually, the relationship between the regression parameters $\lambda_j$ and the noise level is not a simple linear relationship, and the threshold obtained in (10) cannot qualitatively reduce the noise of the noisy image. In practice, the regression parameter $\lambda_j$ is more likely to be a quadratic polynomial of the noise level. Therefore, we propose a quadratic relationship model between the regression parameters $\lambda_j$ and the noise level $\sigma$ in this paper, which is defined as follows:

$$\lambda_j = \frac{k\sigma^2 + b\sigma}{2^{1/2}}$$ (10)

After a simple image denoising experiment, we set three sets of commonly used parameter values for the denoising models corresponding to the three penalty functions, as shown in Table 1.

|        | logarithmic | arctangent | rational |
|--------|-------------|------------|----------|
| $k$    | 0.05        | 0.04       | 0.04     |
| $b$    | 2.5         | 4          | 3.5      |
| $k$    | 0.06        | 0.05       | 0.05     |
| $b$    | 2           | 3.5        | 3.0      |
| $k$    | 0.07        | 1.5        | 3        |
| $b$    | 0.06        | 3          | 2.5      |

3. Experiment and result

We compare the improved penalty functions based improved image denoising algorithm with the penalty functions based original image denoising methods and several widely used denoising methods, including hard and soft methods[21-23]. In subsection 3.1, we introduce the assessment parameters, and in subsection 3.2, we evaluate the improved wavelet denoising algorithm and its competing methods on 4 widely used test images.

![Figure 1. The 4 test images. (a) Lena; (b) Boat; (c) Peppers; (d) Barbara.](image)

3.1. Assessment parameters: PSNR and SSIM

The measure of error is Peak Signal to Noise Ratio [24][25], which is more commonly used in benchmarking algorithms in image processing. The PSNR between two 8-bit gray-scale images is defined to be

$$PSNR = 10\log\left(\frac{255^2}{MSE}\right)$$ (11)

In addition to the above methods, with the deepening of human's understanding of human visual system (HVS), the structural similarity theory (SSIM) evaluation method, which makes use of some
characteristics of human visual system for evaluation, starts to be applied in image quality evaluation to compare the brightness, contrast and structural information of reference images [27].

\[
SSIM(f, g) = \left[ l(f, g) \right]^{\alpha} \left[ c(f, g) \right]^{\beta} \left[ s(f, g) \right]^{\gamma}
\]

(12)

3.2. Results and discussions

We evaluate the improved methods on 4 widely used test images, whose scenes are shown in Fig. 1. Zero mean additive white Gaussian noises with variance \(\sigma^2\) are added to those test images to generate the noisy observations. Due to page limit, we show the results on three noise levels, ranging from low noise level \(\sigma = 10\), to medium noise levels \(\sigma = 50\), and to strong noise level \(\sigma = 100\). The PSNR and the SSIM results by the competing denoising methods are shown in Table 2. The highest PSNR result for each image and on each noise level is highlighted in bold. We have the following observations.

| \(\sigma\) | Image | Parameters | Original | Hard | Soft | Log | Atan | Rat | ILog | IAtan | IRat |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | Lena | PSNR | 28.14 | 29.41 | 33.20 | **33.35** | 33.16 | 33.29 | **33.35** | 33.18 | 33.30 |
| | SSIM | 0.61 | 0.70 | 0.86 | **0.87** | 0.86 | 0.87 | **0.87** | 0.86 | 0.86 |
| | PSNR | 28.11 | 29.05 | 31.03 | 31.34 | 31.13 | 31.11 | **31.50** | 31.13 | 31.34 |
| | SSIM | 0.69 | 0.73 | 0.80 | 0.83 | 0.82 | 0.82 | **0.83** | 0.82 | **0.83** |
| | Peppers | PSNR | 28.16 | 29.18 | 31.10 | 31.39 | 31.18 | 31.40 | **31.53** | 31.18 | 31.35 |
| | SSIM | 0.68 | 0.74 | 0.86 | 0.87 | 0.86 | 0.86 | **0.86** | 0.86 | **0.86** |
| | Boat | PSNR | 28.12 | 29.08 | 30.14 | 31.26 | 31.34 | 31.11 | **31.50** | 31.11 | 31.34 |
| | SSIM | 0.72 | 0.76 | 0.85 | **0.88** | 0.87 | 0.88 | 0.87 | **0.88** |
| | Barbara | PSNR | 28.13 | 29.18 | 31.37 | 31.75 | 31.55 | 31.69 | **31.91** | 31.55 | 31.75 |
| | SSIM | 0.68 | 0.73 | 0.84 | **0.86** | 0.85 | 0.86 | **0.86** | 0.85 | **0.86** |
| AVE | PSNR | 28.13 | 29.18 | 31.37 | 31.75 | 31.55 | 31.69 | **31.91** | 31.55 | 31.75 |
| | SSIM | 0.68 | 0.73 | 0.84 | **0.86** | 0.85 | 0.86 | **0.86** | 0.85 | **0.86** |
| 50 | Lena | PSNR | 14.16 | 16.02 | 23.26 | 24.62 | 24.60 | 24.28 | 26.22 | 26.12 | **26.23** |
| | SSIM | 0.11 | 0.15 | 0.40 | 0.48 | 0.47 | 0.49 | 0.67 | 0.68 | 0.67 |
| | Boat | PSNR | 14.13 | 15.93 | 22.61 | 24.26 | 24.10 | 24.26 | 24.10 | 24.20 |
| | SSIM | 0.15 | 0.20 | 0.41 | 0.49 | 0.47 | 0.57 | 0.57 | 0.57 |
| | Peppers | PSNR | 14.13 | 16.00 | 22.03 | 22.59 | 22.46 | 23.37 | 23.19 | **24.26** | 24.10 |
| | SSIM | 0.17 | 0.22 | 0.45 | 0.53 | 0.51 | 0.49 | 0.57 | 0.57 | **0.57** |
| | Barbara | PSNR | 14.15 | 15.84 | 21.83 | 23.46 | 23.37 | 23.19 | **24.26** | 23.92 | 24.02 |
| | SSIM | 0.19 | 0.23 | 0.43 | 0.49 | 0.47 | 0.57 | 0.56 | 0.56 |
| | Ave | PSNR | 14.14 | 15.95 | 22.43 | 23.24 | 23.13 | 22.98 | **24.09** | 23.92 | 24.02 |
| | SSIM | 0.16 | 0.20 | 0.42 | 0.50 | 0.49 | 0.49 | 0.57 | 0.57 | **0.57** |
| 100 | Lena | PSNR | 8.13 | 10.05 | 17.99 | 19.84 | 19.93 | 19.37 | **23.68** | 23.45 | 23.64 |
| | SSIM | 0.04 | 0.06 | 0.18 | 0.28 | 0.30 | 0.26 | **0.63** | 0.60 | **0.63** |
| | Boat | PSNR | 8.14 | 10.03 | 17.64 | 19.30 | 19.33 | 18.88 | **22.14** | 21.95 | 22.08 |
| | SSIM | 0.05 | 0.10 | 0.20 | 0.28 | 0.29 | 0.26 | **0.50** | 0.47 | **0.50** |
| | Peppers | PSNR | 8.10 | 10.04 | 17.30 | 18.50 | 18.49 | 18.14 | **20.43** | 20.42 | 20.20 |
| | SSIM | 0.06 | 0.12 | 0.23 | 0.31 | 0.33 | 0.30 | **0.56** | 0.54 | 0.54 |
| | Barbara | PSNR | 8.12 | 10.00 | 17.36 | 18.70 | 18.72 | 18.32 | **21.24** | 21.12 | 21.19 |
| | SSIM | 0.07 | 0.12 | 0.22 | 0.28 | 0.29 | 0.26 | **0.49** | 0.47 | **0.49** |
| | Ave | PSNR | 8.12 | 10.03 | 17.57 | 19.09 | 19.12 | 18.68 | **21.87** | 21.74 | 21.78 |
| | SSIM | 0.06 | 0.10 | 0.21 | 0.29 | 0.30 | 0.27 | **0.55** | 0.52 | 0.54 |

First, the wavelet threshold denoising methods using penalty function are obviously superior to the sub-band soft and hard threshold methods. It achieves (2.5dB-8.9dB/0.13-0.30) \((PSNR/SSIM)\) improvement over the sub-band soft threshold method in average and outperforms the sub-band hard threshold method by (0.1dB-0.69dB/0.01-0.08) \((PSNR/SSIM)\) in average on all the four noise levels.
Second, with the noise level increases, the denoising performance of the penalty function threshold method based on the linear regression parameter model decreases. In theory, the logarithmic penalty function should have the strongest induced sparsity, but in this model, as the noise level increases from 50 to 100, the logarithmic penalty function gradually loses its advantage and the $PSNR$ value is no longer the highest. Third, the improved logarithmic, arctangent and rational penalty functions show better performance than the original logarithmic, arctangent and rational penalty functions. And the $PSNR$ values are increased by 0.14dB-2.79dB, 0.01dB-2.62dB, and 0.06dB-3.10dB, respectively, the $SSIM$ values are increased by 0.00-0.26, 0.00-0.22, and 0.00-0.27, respectively. Among the three penalty functions, the threshold function based on the logarithmic penalty function at any noise level can stably exhibit the optimal performance, which is consistent with the theory in [16].

In summary, the improved penalty function wavelet denoising algorithm shows strong denoising ability, can retain more original image features, and has a higher $PSNR$ index.

4. Conclusions
As a significant extension of the convex problem, the parametric penalty function was studied in this paper. We showed that, when the penalty function is applied to the wavelet denoising algorithm, the sparsity of the wavelet coefficients can be well induced. When the noise level is too high, the performance of the penalty function will become less and less optimal. We presented an improved algorithm to solve it. When the relationship between the regularization parameter and the noise level rises from a linear relationship to a quadratic polynomial, a new regularization parameter model is established. We then applied the proposed algorithm to image denoising. The experimental results showed that, compared to the regularization parameter model under linear relations, the regularization parameter model enables an improved denoising algorithm to achieve better performance in almost all cases, especially at higher noise levels. It can be expected that the parametric penalty function will have more successful applications in computer vision problems.

Acknowledgments
The authors gratefully acknowledge the financial supports by the National Key R&D Program of China under grant numbers 2017YFB1201201.

References
[1] Chang, S.G., Yu, B., Vetterli, M. (2000) Adaptive wavelet thresholding for image denoising and compression. IEEE Transactions on Image Processing, 9: 1532-1546.
[2] Donoho, D.L. (1995) De-noising by soft-thresholding. IEEE Transactions on Information Theory, 41: 613-627.
[3] Srivastava, M., Anderson, C.L., Freed, J.H. (2016) A New Wavelet Denoising Method for Selecting Decomposition Levels and Noise Thresholds. IEEE Access, 4: 3862-3877.
[4] Zhang, L., Bao,P., Pan, Q. (2001) Threshold analysis in wavelet-based denoising. Electronics Letters, 37: 1485-1486.
[5] Cetin, A.E., Tofighi, M. (2015) Projection-based wavelet denoising [lecture notes]. IEEE Signal Processing Magazine, 32: 120-124.
[6] Wu, Y.I., Gao, G.J., Cui, C. (2019) Improved Wavelet Denoising by Non-Convex Sparse Regularization Under Double Wavelet Domains. IEEE Access, 7: 30659-30671.
[7] Blumensath, T., Davies, M.E. (2009) Iterative hard thresholding for compressed sensing. Applied and Computational Harmonic Analysis, 27: 265-274.
[8] Chen, S.B., David, L. D., Michael, A. S. (2006) Atomic decomposition by basis pursuit. Society for Industrial and Applied Mathematics,43: 129-159.
[9] Charnigo, R., Sun, J.Y., Muzic, R. (2006) A semi-local paradigm for wavelet denoising. IEEE Transactions on Image Processing, 15: 666-677.
[10] Rajan, S., Wang, S.C., Inkol, R., Joyal, A. (2006) Efficient approximations for the arctangent function. IEEE Signal Processing Magazine, 23: 108-111.
[11] Bayram, İ. (2015) On the Convergence of the Iterative Shrinkage/Thresholding Algorithm with a Weakly Convex Penalty. IEEE Transactions on Signal Processing, 64: 1597-1608.

[12] Zhang, C.H. (2010) Nearly Unbiased Variable Selection Under Minimax Concave Penalty. The Annals of Statistics, 38: 894-942.

[13] Selesnick, I., Farshchian, M. (2017) Sparse Signal Approximation via Nonseparable Regularization. IEEE Transactions on Signal Processing, 65: 2561-2575.

[14] Lanza, A., Morici, S., Sgallari, F. (2016) Convex Image Denoising via Nonconvex Regularization with Parameter Selection. Journal of Mathematical Imaging and Vision, 56: 195-220.

[15] Mohimani, H., Babaie-Zadeh, M., Jutten, C. (2008) A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed $l^0$-Norm. IEEE Transactions on Signal Processing, 57: 289-301.

[16] Selesnick, I.W., Bayram, I. (2014) Sparse Signal Estimation by Maximally Sparse Convex Optimization. IEEE Transactions on Signal Processing, 62: 1078-1092.

[17] Wu, Y.J., Gao, G.J., Cui, C. (2019) Improved Wavelet Denoising by Non-Convex Sparse Regularization Under Double Wavelet Domains. IEEE Access, 7: 30659-30671.

[18] James S. W. (2002) A primer on WAVELETS and Their Scientific Applications. CRC Press, Boca Raton.

[19] Wright, S. J., Nowak, R. D., Figueiredo, M. A. T. (2009) Sparse reconstruction by separable approximation. IEEE Trans on Signal Process, 57: 2479-2493.

[20] Ding, Y., Selesnick, I. W. (2015) Artifact-free wavelet denoising: Non-convex sparse regularization, convex optimization. IEEE Signal Process, 22:1364-1368.

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

[22] Donoho, D.L., Johnstone, I.M. (1995) Adapting to unknown smoothness via wavelet shrinkage. Journal of the American Statistical Assoc, 90: 1200-1224.

[23] Donoho, D.L. (1995) De-noising by soft-thresholding. IEEE Trans Inform Theory, 41: 613-627.

[24] Wang, Z., Bovik, A.C. (2002) Universal image quality index. IEEE Signal Process, Lett.

[25] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process, 13: 600-612.

[26] Chang, S.G., Yu, B., Vetterli, M. (2000) Adaptive wavelet thresholding for image denoising and compression. IEEE Trans Image Processing, 9: 1532-1546.

[27] Moulin, P., Liu, J. (1999) Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. IEEE Trans Inform Theory, 45: 909–919.

[28] Bach, F., Jenatton, R., Mairal, J., Obozinski, G. (2012) Optimization with Sparsity-Inducing Penalties. Found Trends Mach Learn, 4: 1-106.

[29] Blumensath, A. (2012) Accelerated iterative hard thresholding. Signal Process, 92: 752-756.

[30] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J. (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn, 3: 1-122.