An optimization method of signal de-noising in discrete wavelet transform based on generalized cross-validation

Xiaojing Chen, Ke Liu, Peng Ye[*]

College of Physics and Electronic Engineering Information, Key Laboratory of Low-voltage Apparatus Intellectual Technology of Zhejiang, Wenzhou University, Chashan University Town, Wenzhou, Zhejiang Province, 325035, P.R. China

Abstract

A method for automatically selecting the asymptotical optimal parameters is presented for signal de-noising in discrete wavelet transform. The parameters of wavelet de-noising were first encoded. A generalized cross-validation algorithm was then used to select these parameters automatically. The parameters that obtained the smallest generalized cross-validation were asymptotically optimal. Simulation signals with different features and range signal-to-noise ratios were used to demonstrate the optimality of the proposed method. In addition, the Raman spectrum of edible oil and nuclear magnetic resonance spectrum of quinine and Boc-protected proline were employed as real-world data to validate the proposed method. The proposed method achieved superior performances in both real-world data and in artificial simulation.

Keywords: signal de-noising, wavelet transform, generalized cross-validation, parameter optimization

1 Introduction

Obtaining accurate characteristics of instrumental signal is important, however, the required signal is often polluted by noise. To extract reliable information from instrumental signals such as infrared spectra (IR), Raman spectral, unwanted noise should be removed [1, 2]. The effects of noise in spectra can be reduced in several ways. One of the most recent methods is based on wavelet transform (WT). Because of the advantages of the localization of time-frequency characteristics and the sparse representation of signal in the wavelet domain, discrete wavelet transform (DWT) is becoming an increasingly important tool in and compression [3-7]. The most popular tool in wavelet de-noising is the wavelet coefficient shrinkage method known as wavelet threshold de-noising [8-10]. The quality of wavelet threshold de-noising is strongly affected by several essential parameters, including wavelet function, decomposition levels, threshold estimate, and thresholding policy [2]. How to obtain these optimal parameters is an important issue for wavelet de-noising. To overcome this problem, cross-validation (CV) and generalized cross-validation (GCV) have been proposed to optimize the threshold estimate [2, 11-16]. Pasti et al. proposed a method to optimize individually the parameters, including the optimal decomposition level, wavelet function, and threshold estimate [2], Cai et al. employed minimum description length (MDL) algorithm to select the threshold and wavelet functions [17]. However, individual optimal parameters do not guarantee a global optimal when all parameters are considered simultaneously. To the best of our knowledge, no single method can optimize these parameters automatically and simultaneously in DWT threshold de-noising.

The present study obtains the optimal parameters by maximizing DWT threshold de-noising. The DWT threshold de-noising parameters are first encoded using Arabic numerals. The GCV algorithm, which was originally developed to optimize the soft threshold value, is then applied to determine automatically the optimal parameter combinations so that the optimal performance of DWT threshold de-noising can be reached.

2 Theory

2.1 DISCRETE WAVELET TRANSFORM (DWT)

Discrete wavelet transform is a linear transform, and the wavelet transform of a discrete signal f can be described as the follows:

$$w = Wf , \tag{1}$$

where W is an orthonormal matrix represented as wavelet basis or wavelet filter coefficients, and w I is the wavelet transform coefficients. f is decomposed by the filterbank with a lowpass filter and highpass filter of W into set of wavelet coefficients w:

$$w = [cD_1, cD_2, ..., cD_J, cA_J] , \tag{2}$$

where J is decomposition level, and $cD_1, cD_2, ..., cD_J$ represent the detail information of signal f and cA_J represents the approximation information. This transform localizes the most important spatial and frequent characteristics of f in a limited number of wavelet coefficients since the wavelet basis can be derived from a common function called

---

[*]*Corresponding author* e-mail: yep@wzu.edu.cn

mother wavelet, which has two operations: translation and dilation.

Signal reconstruction can be processed by the same highpass filter and lowpass filter in the W can be described by the simple equation:

$$f = W^t w , \qquad (3)$$

where W^t is inverse wavelet transform. In practice, a fast algorithm developed by Mallat [18] is commonly used for performing the transformation.

2.2 DWT THRESHOLD DENOISING

An input instrumental signal can be represented as the sum of two components:

$$y = f + \varepsilon , \qquad (4)$$

where f represents the ideal signal and ε represents the noise. A DWT yields the same situation in terms of wavelet coefficients:

$$w = v + \omega . \qquad (5)$$

The vector $v = Wf$ contains the wavelet coefficients of the original instrumental signal. Here, $\omega = W\varepsilon$ is the noise coefficient and $w = Wy$ is the noisy signal coefficient. To obtain the ideal signal f, v and ω are separated from one another using a comparison of their wavelet coefficients. The simplified procedure for DWT threshold de-noising can be described as follows:

1) A wavelet transform is applied to signal y to obtain the wavelet coefficient w.

2) The wavelet coefficient ω is removed to obtain the estimated wavelet coefficient f_d.

3) The inverse discrete wavelet transform is applied to f_d to obtain the de-noised signal y_d.

One of the key issues of DWT threshold de-noising is how to estimate ω (i.e., how to distinguish ω from v). Many possible approaches can be used to estimate the noise level, a systematic analysis of the performance of these approaches can be found in [1, 19]. The universal threshold has the following format:

$$\text{TUV} = \sigma\sqrt{2\ln N} , \qquad (6)$$

where N is the length of signal y, σ is the standard deviation of the noise, and estimated from the median of the detail coefficients at the first level of signal decomposition.

$$\sigma = \text{median}\left|(\text{de tai1})\right| / 0.674 . \qquad (7)$$

Once the threshold value has been calculated, the key question is how to obtain f_d. In general, DWT threshold de-noising methods use two different policies: hard and soft thresholding. The hard thresholding policy simply sets all the wavelet coefficients below a certain threshold th to zero:

$$\begin{cases} w_{ij}^t = w_{ij} & if \ |w_{ij}| > th \\ w_{ij}^t = 0 & if \ |w_{ij}| < th \end{cases} . \qquad (8)$$

In soft thresholding, the values of the wavelet coefficients are shrunk by a certain threshold if they are above a certain threshold th :

$$\begin{cases} w_{ij}^t = \text{sgn}(w_{ij})(|w_{ij}| - th) & if \ |w_{ij}| \geq th \\ w_{ij}^t = 0 & if \ |w_{ij}| < th \end{cases} , \qquad (9)$$

where sgn is the sign function that returns the sign of the wavelet coefficient w_{ij} .

2.3 GENERALIZED CROSS-VALIDATION (GVC)

The goal of signal de-noising is to minimize the difference between the de-noised and the ideal (i.e., noiseless) signal, thereby minimizing the mean square error (MSE), as shown below:

$$\text{MSE} = \frac{1}{N}\|f_d - f\| = \frac{1}{N}\sum_{i=1}^{N}\left(f_d(i) - f(i)\right)^2 , \qquad (10)$$

where f_d is the de-noised signal, f is the ideal signal. However, normally, the ideal signal is not known. Thus, MSE cannot be obtained.

The GVC theory was developed to estimate the best threshold value to optimize DWT threshold de-noising using the soft thresholding policy. GVC defines the risk estimate function as follows:

$$GCV(\xi) = N\|w - w_\xi\|^2 / N_0^2 , \qquad (11)$$

where N_0 is the number of the coefficients replaced by zero, N is the total number of the wavelet coefficients, w_ξ is the modified wavelet coefficient after applying a threshold ξ, and w is the wavelet coefficient of the original noisy signal. [12] proved that the threshold ξ that results in the smallest GCV is asymptotically optimal under certain conditions. Moreover, [20] improved the GCV algorithm for the hard thresholding policy. The detailed description of the GVC theory can be found in the document [20]. In general, when GCV is smallest, MSE is asymptotically at minimum, which means searched parameter for de-noise might be optimal, and might also rank among the top 10%, and implies that the de-noising result is asymptotically optimal. Moreover, GVC only depends on input and output data, which are of vital importance in practice. Therefore, GCV is able to measure MSE when the shape of the signal is unknown or the noise energy is difficult to estimate.

3 Experimental and calculations

3.1 ENCODING OF DWT THRESHOLD DENIOSING PARAMERERS

Many types of wavelet functions exist, and some wavelet families contain wavelets of different orders. The Db, Bior, Sym, and Rbio wavelets which are the different filters have been proven effective in signal processing [21]; however, the present study is limited only to the daubechies, symmlet, bior, and rbio families of filters. A total of forty-eight filters

were investigated in the present study. To obtain the minimum GCV, these wavelet functions need to be encoded as variables denoted as Arabic numerals. For example, "1" denotes the wavelet function "sym1," "11" denotes the wavelet function "db1," and "48" denotes the wavelet function "dmey." At the same time, other parameters, such as decomposition levels, threshold policy, and threshold estimation, are also encoded as Arabic numerals. The definitions and the encoding of these parameters are shown in Table 1.

TABLE 1 Definition and Encoding of wavelet de-noising parameters

Parameters description	Definition domain	Remark
Wavelet function	$[1, N_{wavelet\ name}]^{[1]}$	$N_{wavelet\ name}$ is the number of wavelet function in definition domain. Here, $N_{wavelet\ name}$ is 48.
Decomposition scales	$[1, N_{max}]$	N_{max} is the maximum of decomposition scales. Here, N_{max} is 10.
Threshold estimation	[5Tuv, 8Tuv]	Tuv is universal threshold.
Thresholding policy	[1, 2]	Hard or soft thresholding.

[1] Wavelet filters from 1 to 48: sym1, sym2, sym3, sym4, sym5, sym6, sym7, sym8, sym9, sym10, db1, db2, db3, db4, db5, db6, db7, db8, db9, db10, db12, db15, db18, db20, db30, bior1.1, bior1.3, bior2.2, bior2.4, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8, rbio1.1, rbio1.5, rbio2.2, rbio3.3, rbio3.5, rbio3.7, rbio3.9, rbio4.4, rbio5.5, rbio6.8, Dmey;

3.2 SIMULATION SIGNALS AND CALCULATIONS

Different simulation signals were considered in order to evaluate the performance of the proposed method. The simulation signals of Blocks, Bumps, Heavysine and Doppler signal in the wavelet tool are employed in present study. The four signals contain abundant frequency components, many peak information and lots of catastrophe parts, so they are always used as the typical signals for comparing

the quality of de-noising algorithms. The four original signals and noisy signals are showed in Figure 1.

In this study, the Wavelet Toolbox 3.0 (The Math Works, Natick, USA) was used, and the Pieflab Toolbox for GCV algorithm can be obtained from http://homepages.ulb.ac.be/~majansen/software/ thresh -lab.html, and all the calculations were performed on the platform of MATLAB 7.6 (The Math Works, Natick, USA).
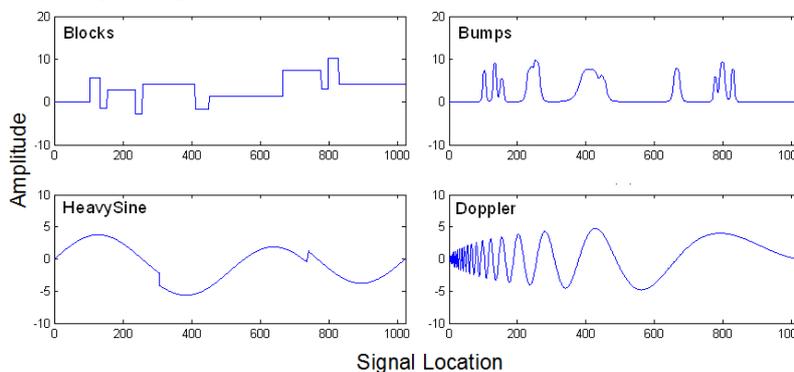

FIGURE 1 Four simulation signals

4 Results and discussions

4.1 SIMULATION DATA

In this study, the definition domain of the threshold was from 5Tuv to 8Tuv with the interval of 1 unit. As a result a total of 40 definition points are available.

As mentioned above, when the minimum GCV value is achieved, MSE is also near to the optimal value under certain conditions. This means that the parameter combinations of DWT threshold de-noising corresponding to the minimum GCV are also near to optimal. For real-world signals, MSE cannot be estimated, hence, rather than

MSE, the proposed GCV approach was adopted to optimize the parameter combinations.

Random noise was added 100 times to each type of simulation signal with the signal-to-noise ratio (SNR) ranging from 16.34 to 21.18 (Level 1) to validate whether the parameter combinations corresponding to the minimum GCV (or optimal GCV) are optimal. Each type of simulation signal can obtain 100 noisy signals. Here, only Doppler noisy signals are listed in Figure 2a (Level1). The MSE value corresponding to the minimum GCV (GCV-MSE) can be obtained from 38,400 (48 * 10 * 40 * 2 parameter combinations) different MSE values. Observing the arrangement locations of GCV-MSE in all 38,400 MSE

values is a simple way to determine whether GCV-MSE is minimum. When GCV-MSE is minimum (or ranks first), and imply the obtained parameter combination is optimal. Similarly, if GCV-MSE is near to minimum (or ranks in the top ten or top one hundred), the searched parameter combination is asymptotical optimal. The GCV-MSE concrete arrangement distribution of 100 noisy signals of each type of simulation signals is shown in Figure 3a. In Figure 3a, the abscissa values for Blocks, Bumps, Heavysine, and Doppler signals are 1–100, 100–200, 201–300, and 301–400, respectively. The small circles in Figure 3a represent the arrangement position of each noisy signal. Each discrete point is connected with solid lines for easy comparison of several different signals.
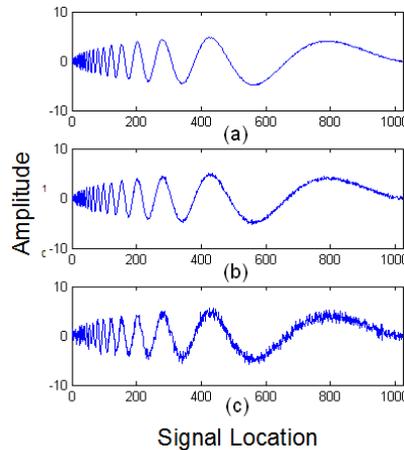

FIGURE 2 Noisy Doppler signals with different SNR level: (a) Level 1 (b) Level 2 (c) Level 3
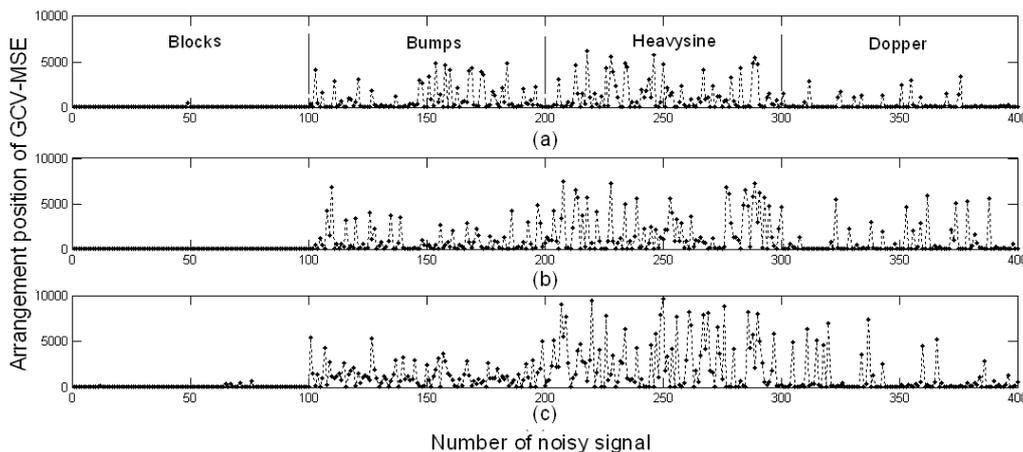

FIGURE 3 GCV-MSE Arrangement locations of four simulation signals with different SNR, 1-100 is the arrangement locations of GCV-MSE for the Blocks signals, 100-200 are for Bumps signals, 201-300 are for Heavysine signals, 301-400 are for Dopler signals:
(a) level 1 SNR (b) level 2 SNR (c) level 3 SNR

As seen in Figure 3a, the arrangement locations of GCV-MSE is relatively small for signals with the SNR level 1, and most of them are lower than 5,000, which indicates that the identified parameter combination is working well. A comparison of the optimization results of four signals shows the results of the Block signal sequence 1–100, as shown in the Figure 3a to be the best. This result can be explained by the GCV algorithm principle. Because the GCV algorithm is more suitable for regular signals, it is more appropriately applied in signals with less frequency information. The mean and standard deviation of GCV-MSE arrangement positions listed in Table 2 further validate the optimization effect of the proposed method. As shown in Table 2, the mean of the arrangement locations of the GCV-MSEs is top-ranked (corresponding to 38,400 combinations), indicating that optimization results are very satisfactory. Compared with these results, the mean of Heavysine is the largest, which indicates that the optimization result is the worst among the four simulation signals. Moreover, the standard deviations of the Blocks signals are smaller than that of the other three signals, which indicate that the proposed method is relatively stable for Blocks signals.

TABLE 2 Mean and standard deviations of GCV-MSE arrangement locations of four simulation signals

Signals SNR level	Blocks		Bumps		Heavysine		Doppler	
	Mean	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*
Level 1	21	48	876	1327	1249	1651	285	665
Level 2	17	16	847	1297	1947	2215	580	1339
Level 3	38	92	1239	1176	2729	2840	690	1586

TABLE 3 Mean and standard deviations of MDL-MSE arrangement locations of four simulation signals

Signals SNR level	Blocks		Bumps		Heavysine		Doppler	
	Mean	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*	*Mean*	*Sd*
Level 1	71	74	2007	981	3897	1895	1131	692

The MDL algorithm is free from any parameter setting or any subjective judgment when used for data compression and de-noising [17]. In the present study, the MDL algorithm can also be used in the parameter optimization of the DWT algorithm, with the coding and setting of the parameters the same as before. The optimization result shown in Table 3 also shows a comparison of the results of GCV algorithm of the noise level 1. Based on the Table 3, the MDL algorithm is inferior to the GCV algorithm, which is consistent with the result proposed by the document [2]. Moreover, MDL can only be used for optimization of hard threshold values. Hence, the following studies are all based on GCV algorithm.

To investigate the influence of adding signals with different noise levels on the proposed method, 100 noisy signals with SNRs ranging from 11.17 to 16.22 (Level 2) and from 8.82 to 11.37 (Level 3) are employed. The noisy signals with different SNRs are shown in Figure 2b and 2c, along with a list of the Doppler noisy signals. Similarly, the GCV-MSE specific arrangement distributions of the SNR levels 2 and 3 are shown in Figures 3b and 3c. Clearly, the performance of the proposed method is influenced by different levels of noises. Based on Figure 3, the Heavysine and Doppler signals are influenced the most, followed by the Bumps signal, whereas little influence is generated on the Blocks signal. These results indicate that the proposed method is suitable for several regular signals, further verifying that the GCV algorithm is more suitable for regular signals. In Figure 3b, most of the GCV-MSE arrangement locations are within 5,000, and only a few Bumps and Heavysine signals exceed 5,000, and the GCV-MSE arrangement locations greater than 5,000 are slightly increased for the SNR level 3, especially for Heavysine signal, as seen from Figure 3c. However, compared with the 38,400 MSE values, the arrangement locations are still placed at the front location. The mean and standard deviations of the arrangement positions of 100 GCV-MSE of the four types of signal with SNRs at Levels 2 and 3 are listed.

As seen from Table 2, the means of all the arrangement locations are small, which indicates that the parameter optimization results are satisfactory. The Heavysine and Doppler signal are greatly affected by noise on the aspect of standard deviation values, whereas noise has little influence on Blocks signals.

At the same time, the specific numbers of different arrangement scopes have been summarized out, as shown in the Table 4. The optimization results of the parameters of three types of noises which is shown in the Table 4 indicate that the numbers of those whose arrangement locations are below 384 are 292, 269 and 227 respectively, which means that there are probability values of 73.00%, 67.25 and 56.75% to select the optimal combinations of top 1% correspondingly. Meanwhile, this also means that there are probability values of 94.50%, 91.75% and 89.25% for the optimal combinations of top 10% to be selected respectively. This result is very satisfactory for such a huge combination of 38400.

In order to investigate the relation between the GCV value and the corresponding MSE values, each GCV value of the 38,400 different parameters and the corresponding MSE values should be listed. However, illustrating such massive data would be impossible because of page limitations. Hence, only 200 data were selected randomly from Block and Heavysine signals, their GCV values and the corresponding MSE values are reported in Figure 4. As one can see from Figure 4a, small GCV is not always corresponds to the small MSE, however, in the region of GCV smaller than 0.5, the trends of GCV and MSE are similar, based on Figure 4b, it was found that the smaller GCV and MSE correspond to same data point (parameter combination). Similar situation is also happened in the Heavysine signal; the detailed results are showed in Figures 4c and 4d.

TABLE 4 Numbers of arrangement location

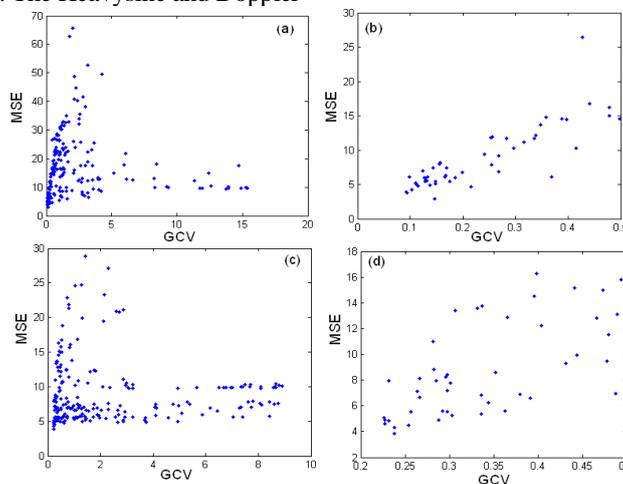| SNR level	Range		
	<384	*<1920*	*<3840*
Level 1	292	356	378
Level 2	269	337	367
Level 4	227	321	357

FIGURE 4 Relations between GCV and MSE of block and heavysine signals

4.2 REAL-WORLD DATA

The real-world data used in the present study is the Raman spectrum of edible oil (mainly composed of protein and fat). The T64000 Raman spectrometer used was produced by the French HORIBA Jobin Yvon Company. All Raman tests were conducted at room temperature. The excitation light source used was an argon ion laser with a wavelength of 514.5 nm, laser power of 120 mw, and scanning wavenumber range of 500–3,500 cm^{-1}. The experimental light path was backscattered. Laser-formed light spots were shown to have a diameter of approximately 1 um on the sample surface after being focused 100 times by the object lens. The original Raman spectrum in Figure 5a shows four peaks with different widths; the figure also shows the height and the area of the spectra peaks. In general, the shapes of the peaks, including width, height, and area should be maintained in the de-noising process to avoid influence on the analysis results. This makes signal de-noising very challenging.

Using the proposed GCV methodology, the optimal parameter combinations were obtained as 8, 5, 4, and 1. Correspondingly, wavelet function of sym8, decomposition level of 5, a 4/5 Tuv threshold and hard threshold were employed. De-noising result in Figure 5b shows that the shapes of several peak values to be well kept, whereas most of the noises were eliminated. The attributes of the peak, including position, height, width, and area, are the main indices that define the components in the Raman spectrum analysis. The changes of the attributes of the peak before and after de-noising are shown in Table 5. The position, width, height, and area of the peak showed no prominent changes after the noise of the original spectrum was eliminated.
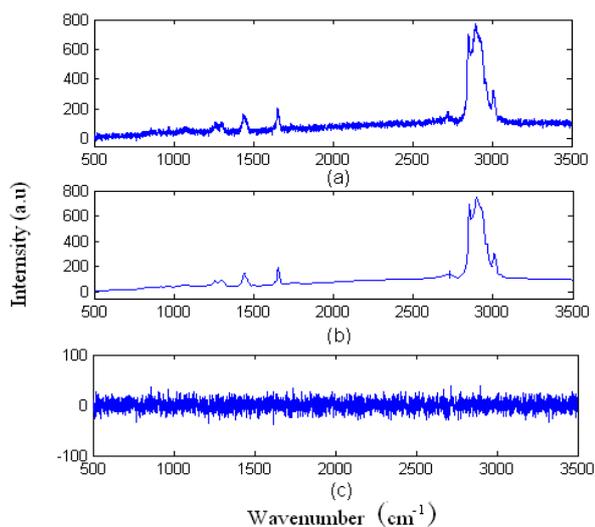

FIGURE 5 Raman spectrum of edible oil: (a) noisy spectrum (b) denoised spectrum (c) noisy signals
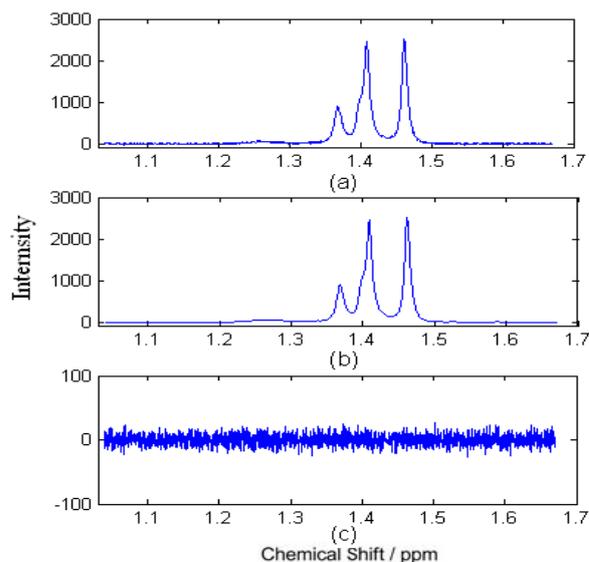

FIGURE 6 NMR spectrum of mixtures of quinine and Boc-protected proline: (a) noisy spectrum (b) denoised spectrum (c) noisy signals

Another real-world data is nuclear magnetic resonance (NMR) spectrum in Figure 6a, which was obtained 16:84 mol ratio mixtures of quinine and Boc-protected proline with enantiomeric excess of 10% in CDCl3. The NMR spectra were recorded after several minutes of thermal equilibration time. NMR spectrum was recorded on a Bruker Avance 500 MHz spectrometer. Spectrum was recorded using 16 scans at 298K. A full spectrum of sample was recorded referenced to TMS. An exponential window function with a line-broadening factor of 1Hz was applied to the FID before Fourier transformation. The 1H-NMR spectrum was phased and baseline-corrected using Topspin 2.1(Bruker) and was automatically reduced by using the AMIX (Bruker GmbH, Germany) software package to continuous integral segments of equal width of 0.004ppm corresponding to the chemical shift range 1H, δ 6.9–8.8 after removing the solvent resonance region (δ 7.2-7.3). Similar to the situation of Raman spectrum, to de-noise NMR spectrum is a very challenging thing because it requires that the shape of the peak should remain unchanged as much as possible. Since several peaks of this NMR spectrum are connected together, there is impossible to calculate the area and width of each peak. Therefore, we give the height of the peak and total area and width of three peaks prior to de-noising here. Using the proposed GCV methodology, the obtained optimal parameter combination is 4, 5, 4 and 1, correspondingly, parameters of wavelet function is sym8, decomposition level is 5, and 4/ 5Tuv threshold and hard threshold are employed. De-noising results are shown in the Figure 6b. The changes of the attributes of the peak before and after de-noising are shown in Table 6, as one can see from the Table 6, the spectral profile after de-noising, the height of the peak changes relatively less, and the shape of peaks almost remain unchanged, indicating that the de-noising effect is relatively satisfactory. Hence, it can be concluded that the parameter combination thus generated is relatively ideal.

TABLE 5 Comparison of attribute of peak before and after de-noising for Raman spectrum

Peak	Original				Denoised			
	Position	*Width*	*Height*	*Area*	*Position*	*Width*	*Height*	*Area*
1	1148	102	73	7079	1147	104	73	7076
2	1346	46	124	7790.	1345	51	121	7788
3	1652	23	174	6456.	1652	23	173	6451
4	3421	153	668	120529	3422	151	668	120525

TABLE 6 Comparison of attribute of peak before and after de-noising for NMR spectrum

	Original Peak			Denoised Peak		
	1	*2*	*3*	*2*	*2*	*3*
Position	1367	1409	1461	1367	1409	1461
Height	903.5	2446.3	2516.6.	900.5	2445.8	2512.6.
Width		403			404	
Area		310320			310350	

5 Conclusions

The present paper proposed an algorithm that automatically selects DWT threshold de-noising parameters based on the GCV algorithm. Parameter optimization includes wavelet function, decomposition level, threshold estimation, and threshold policy. Four simulation datasets and real-world data of Raman spectral and NMR signals were used for validating the proposed algorithm. The results show that the identified minimum GCV value produces a better MSE value. Therefore, DWT threshold de-noising can be optimized automatically through reasonable parameter encoding based on GCV value. At the same time, the optimization ability of the proposed method slightly degrades with the increase of the magnitude of added noise.

Acknowledgements

References

[1] Alsberg BK, Woodward AM, Winson MK, Rowland J, Kell D B 1997 *Analyst* **122**(7) 645-52

[2] Pasti L, Walczak B, Massart DL, Reschiglian P 1999 *Chemometrics and Intelligent Laboratory Systems* **48**(1) 21-34

[3] Alsberg BK, Woodward AM, Kell DB 1997 *Chemometrics and Intelligent Laboratory Systems* **37**(2) 215-39

[4] Komsta L 2009 *Journal of Chromatography A* **1216**(12) 2548-53

[5] Chen X, Wu D, He Y, Liu S 2009 *Analytica Chimica Acta* **638**(1) 16-22

[6] Zhu D, Ji B, Meng C, Shi B, Tu Z, Qing Z 2007 *Journal of Agricultural and Food Chemistry* **55**(14) 5423-8

[7] Mittermayr CR, Nikolov SG, Hutter H, Grasserbauer M 1996 *Chemometrics and Intelligent Laboratory Systems* **34**(2) 187-202

[8] Donoho DL 1995 *IEEE Transactions on Information Theory* **41**(3) 613-27

[9] Donoho DL, Johnstone IM 1994 *Biometrika* **81**(3) 425-55

[10] Donoho DL, Johnstone IM 1995 *Journal of the American Statistical Association* **90**(432) 1200-24

[11] Nason GP 1996 *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(2) 463-79

[12] Jansen M, Malfait M, Bulthell A 1997 *Signal Processing* **56**(1) 33-44

[13] Jansen M, Bulthell A 1999 *IEEE Transactions on Image Processing* **8**(7) 947-53

[14] Jansen M, Uytterhoeven G, Bulthell A 1999 *Medical Physics* **26**(4) 622-30

[15] Weyrich N, Warhola GT 1998 *IEEE Transactions on Image Processing* **7**(7) 82-90

[16] Tiwari AK, Shukla KK 2004 *Digital Signal Processing* **14**(2) 38-157

[17] Cai C, Harrington PB 1998 *Journal of Chemical Information and Computer Sciences* **38**(6) 1161-70

[18] Mallat SG 1989 *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7) 674-93

[19] Barclay VJ, Bonner RF, Hamilton IP 1997 *Analytical Chemistry* **69**(1) 78–90

[20] Jansen M 2001 *Noise Reduction by Wavelet Thresholding Springer Press New York*

[21] Singh BN, Tiwari AK 2006 *Digital Signal Processing* **16**(3) 275-87

Authors

Xiaojing Chen, born on January 3, 1978, Wenzhou, Zhejiang Province, P.R. China

Current position, grades: associate professor in Wenzhou University, China.
University study: PhD degree was earned in major of optical detection, Xiamen University in 2009.
Research activities: spectra analysis; pattern recognition; digital signal processing.

Ke Liu, born on July 15, 1989, Changsha, Hunan Province, P.R. China

Current position, grades: Graduate student in Wenzhou University,China.
University study: BsC in electronic and information engineering, Hunan University of Science and Engineering in 2012.
Research activities: pattern recognition, digital signal processing.

Peng Ye, born on September 19, 1970, Wenzhou, Zhejiang Province, P.R. China

Current position, grades: lecturer in Wenzhou University, China.
University study: BsC in applied physics, Fudan University, in 1994.
Research activities: optical analysis; electronic engineering.