A New Signal De-Noise Method Using Adaptive Wavelet Threshold based on PSO Algorithm and Kurtosis Measuring for Residual Noise

Sameer AbdulKadhim
Email: samirabdcathem@yahoo.com
Electrical Engineering Department, College of Engineering, University of Babylon

Ahmed Hussein Shatti
Email: eng_ahmed.shatti2014@yahoo.com

Abstract
The signal de-noising based on waveletthresholding is subjected to the value of threshold and how the way selection for it. This made a threshold value acts as an oracle which distinguishes between noise and signal. To date, there have been several methods developed to predict the value of threshold depending on statistical calculations for the noisy signal assuming that there is some prior knowledge for original signal and noise distributions. In fact, in any practical issues, only the observed noisy signal that we hold. Therefore, in this work, an intelligent model is developed to estimate the value of threshold without any prior knowledge for these distributions. This is done by applying the Particle Swarm Optimization (PSO) algorithm for kurtosis measuring of the residual noise signal to find an optimum threshold value at which the kurtosis function be maximum. These residual noise signals can be estimated by applying an inverse threshold function to the detail coefficients of the DWT. This model has been validated by comparison the results with the statistical models and shows strong agreement for the obtained threshold parameter.

Keywords: Signal de-noising, inverse threshold, residual noise, PSO, kurtosis.

1-Introduction
In recent years signal de-noising became one of the most important issues in the area of digital signal processing. The corruptions of signal by noise are common during its processing, compression, transmission, and reproduction. Signal denoising algorithm tries to recover the cleaned version of the signal from its noisy one by removing the noise and retaining the maximum possible signal information.

The capability of the wavelet transform in precisely separating the high and low frequency components in the signal made it works as a dominant technology in the fields of signal de-noising. Because the noise are frequently localized at high frequency components in the signal, therefore, it becomes very useful to use a wavelet transform for decomposing the signal into its different frequency components and...
then get rid of the noise by thresholding it by a suitable threshold value. It should be noted here ifa large value of the threshold used, this will lead to destroy the signal data, while the small value of the threshold retains the noisy data. However, the threshold process modifies the signal data as the per selected threshold value. Therefore, the value of threshold plays an important part in signal de-noising and should be carefully selected (KS Thyagarajan, 2006, Gao, R.X. and Yan, R., 2011).

In the works like (S. Grace Chang et.al., 2000) and (Pankaj Hedao and Swati S Godbole, 2011) the threshold selection has been derived in a Bayesian method using generalized Gaussian distribution (GGD) as a probabilistic model of the signal wavelet coefficients. In last few years, optimization algorithms have been made a revolutionary development in the threshold selection issue. In (Xing et al., Siwei et al., 2014), (Dinesh et al., 2015) and (V. Gopinath, et al., 2014) MSE or signal statistics is used as a fitness function for the optimization algorithms that made these work need a prior knowledge for signal statistical properties.

In this paper, we used PSO algorithm as an optimization technique that depends on a novel criteria for fitness function that rely on the kurtosis measuring for the estimated residual noise signal. Inverse threshold function was innovated to estimate the residual noise from the detail coefficients of the DWT of the noisy signal. Our proposed algorithm suppose that there is a single value for the threshold called optimum threshold that maximizes kurtosis value of the residual noise which is then discovered by PSO algorithm. The robust points in this criteria that it’s no need for any prior knowledge for original statistical properties of the noisy signal. In other words, our proposed algorithm can be used for any signal at any SNR level.

This paper organized as follows: Section (2) and (3) survey the methods related to the traditional wavelet de-noising and traditional threshold selection respectively. Section (4) outlines our proposed algorithm. In section (5) the PSO algorithm is presented. Section (6) discusses the results and performance analysis of the proposed model corresponding to other methods. Finally Section (7) concludes our paper.

2- Traditional Wavelet De-Noising Methods

A noisy signal with additive noise is modeled by:

\[ n_{\text{Sig}} = \text{Sig} + \text{Noise} \] (1)

Where \( n_{\text{Sig}} \) denotes to the observed noisy signal, \text{Sig} is the unknown original signal and \text{Noise} is an independent identically distributed (\text{iid}) random Gaussian noise with zero mean and finite variance.\(^2\). The goal is to remove the noise, or “de-noise” the observed \{ \( n_{\text{Sig}} \) \}, to obtain an estimated \{ \( e_{\text{Sig}} \) \} of the original \{ \text{Sig} \} with minimum mean square error (MSE):

\[ \text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (S_{\text{Sig}} - e_{\text{Sig}})^2 \] (2)

Where \( N \) is the signal length (should be integer power of 2) (S. Grace Chang et al., 2000).

Many de-noising techniques are proposed to overcome this problem, but the most powerful one that using a wavelet transform. Wavelet transform is a well-known tool for signal analysis. It can decompose the signal into many segments which belong to different frequency components. This is accomplished by comparing the signal with a group of wavelet basis functions and then looking for their similarities in frequency contents (Gao et al., Yan et al., 2011).

Let \( W \) denote to the orthogonal Discrete Wavelet Transform (DWT) matrix, then the wavelet coefficients is:

\[ W \times n_{\text{Sig}} = W \times \text{Sig} + W \times \text{Noise} \]

\[ W_{\text{Sig}} = W_{\text{Sig}} + W_{\text{Noise}} \] (3)
Where, $W_{\text{Noise}}$ is also \((iid)\) noise since the transformation is orthogonal (S. Grace Chang et al., 2000).

Practically, the discrete wavelet transform is implemented by using a perfect reconstruction filter bank each of which represent an orthogonal wavelet basis function. The result of this process is a multilevel decomposition, in which the signal is divided at each level into sub bands called \textit{approximation} and \textit{detail} coefficients as shown in Figure (1). If we denote the detail coefficients of the multilevel transform by $cD$ then $cD = [cD_j, \ldots, cD_k, cD_2, cD_1]$ where $k$ is the scalar, with $j$ being the largest (or coarsest) scale in the decomposition, each subband scale $k$ has a length equal to $\left(\frac{N}{2^k}\right)$. The subband $cA_j$ denotes to the approximate coefficients ($cA$) and represents the low-resolution details of the signal (KS Thyagarajan, 2006).

![Figure 1: a) Subbands of 3 levels wavelet decomposition. b) Vector representation of the decomposed signal.](image)

The traditional way in wavelet de-noising method start with trimming each coefficient from the detail subbands ($cD$) with a certain threshold to obtain a threshold version of detail sub bands ($Z$) as shown in Fig. (2). Then ($Z$) reconstructed with the approximation coefficients ($cA$) to produce the estimated or de-noised signal where:

$$
e\text{Sig} = W^{-1} \times [cA_l, Z]$$

Where, $W^{-1}$: referred to the Inverse Discrete Wavelet Transform (IDWT) operator (KS Thyagarajan, 2006).

There are two main threshold functions that frequently used. The \textit{soft-threshold function} (also called the shrinkage function), which is defined as:

$$Z = \psi(cD, T) = \text{sign}(cD) \times \max\{(|cD| - T), 0\}$$

It takes the argument and shrinks it toward zero by the threshold $T$. The other popular alternative is the \textit{hard-threshold function}, which is defined by:

$$Z = \psi(cD, T) = cD \times \mathbb{1}(|cD| \geq T)$$

Where $\mathbb{1}()$ is a logic function (0 or 1), this function keeps the input if its value larger than threshold $T$ otherwise, set it to zero (S. Grace Chang et al, 2000).
Figure 2: Traditional threshold wavelet de-noising model.

3-Traditional Threshold Selection Methods

The main difference between all existing wavelet de-noising methods is how to choose the way in which the threshold value is selected. There are many threshold selection methods that have been developed over the years such as VisuShrink, SureShrink, and BayesShrink. The VisuShrink threshold is evaluated by the following expression:

$$T_{Visu} = \sigma_N \sqrt{2 \log(L)}$$  \hspace{1cm} (7)

Where $\sigma_N$ represents a noise variance and $L$ is a length of signal. This method results in an estimate asymptotically optimal in the minimax sense (minimizing the maximum error over all possible L-sample signals) (S. Grace Chang et al., 2000).

Another notable threshold is Sure-Shrink threshold which is defined by:

$$T_{Sure} = \min\{t_j, \sigma \sqrt{2 \log(L)}\}$$  \hspace{1cm} (8)

Where $t_j$ represents the threshold value at $J$th decomposition level in wavelet domain (Mantosh Biswas and HariOm, 2013).

One of the most popular methods namely, BayesShrink was proposed by (S. Grace Chang et al., 2000) in which the threshold has been derived from Bayesian method. This method has a better performance than the Sure-Shrink in terms of mean square error (MSE). The BayesShrink threshold for every subband is given by:

$$T_{Bayes} = \frac{\sigma^2}{\sqrt{\sigma_{Sig}^2}}$$  \hspace{1cm} (9)

Where $\sigma^2$ noise variance and $\sigma_{Sig}^2$ is the variance of original signal.

4-Proposed Algorithm

Through section (3) all these methods of signal de-noising assume that there is some priori knowledge for signal and noise distributions with known parameters to determine the value of threshold. It is known that in any practical issues, only the noisy signal that observed is determined, therefore, in order to propose a new and effective wavelet de-noising method without depending on this priority of knowledge, in this paper we proposed a model (Figure (3)) that firstly uses the Kurtosis statistic of the residual noises signal to discover the optimum value for threshold at which the Kurtosis function becomes maximum, and then uses the PSO algorithm to reach this value after some iterations.
The proposed denoising model based on PSO algorithm and Kurtosis measuring for residual noise.

The normalized Kurtosis function for any random variable $x$ is defined as:

$$kurt(x) = \frac{E((x-m_x)^4)}{(E((x-m_x)^2))^2} - 3$$  \hspace{1cm} (10)

Where: $E(x)$ is the expected value of $x$. The kurtosis function provides an effective means to probe the statistical properties of random variables. For instance, if $x$ is a Gaussian distribution vector, its kurtosis is always approach to zero (Andrzej and AMARI, 2002).

To estimate the noise added to the signal, the algorithm starts with applying DWT to noisy signal ($nSig$) to decompose it into approximation and detail coefficients. Then a new function is innovated to extract this noise (residual noise $R$) from the detail coefficients. This proposed function is nominated as inverse threshold function and works to shrink the input by $T$ if its absolute value smaller than $2T$, otherwise, set it to $T$.

$$R = \eta(cD, T)$$  \hspace{1cm} (11)

However, to improve the performance of our proposed algorithm, successive approximation techniques can be used. The successive approximation method uses a sequence of thresholds $[T_j, \ldots, T_k, \ldots, T_2, T_1]$ for each sub band and the detail coefficients $[cD_1, \ldots, cD_k, \ldots, cD_2, cD_1]$. Usually, threshold values are halved successively as follows $T_{k} = \frac{T_{k-1}}{2}$ (KSThyagarajan, 2006). But in our problem this is not strictly true, since the amount of noise in each sub band is random. Hence, the accepted values are: $T_1 > T_2 > \ldots > T_j$.

**5-Particle Swarm Optimization (PSO)**

The PSO algorithm is an evolutionary computation algorithm which has been developed by Eberhard and Kennedy in 1995. It simulates the social behavior of bird flocking or fish schooling while searching for the food. In PSO, each particle in...
the swarm represents a possible solution of the optimization problem, which is defined by its fitness function. At each iteration, a new location for the particles is evaluated based on its last position and velocity. In other words, every particle has a one chance to move for each iteration by a magnitude of velocity. So if the velocity is very high the particle will take bigger steps, and if the velocity is very small the particle will move in small steps making the convergence very slowly. Initially, the PSO algorithm deploys the particles randomly within the search space, and then it simply uses the objective function to estimate the fitness of each particle. Therefore, every particle will have a position, fitness value, and velocity. In this case, the best fitness value defined as a best particle or a best individual solution. Finally, the PSO algorithm estimates the global best solution (particle position which gives maximum or minimum fitness value among all particles in the population).

The following equations depict the concept of the standard PSO algorithm which uses both the current global best and the current individual best to reach the desired value after some iteration $t$:

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (12)$$

$$v_i^{t+1} = w v_i^t + c_1 \varepsilon_1 (x_i^* - x_i^t) + c_2 \varepsilon_2 (g_i^* - x_i^t) \quad (13)$$

Where $w$ is the inertia weight, $\varepsilon_1$ and $\varepsilon_2$ are uniform random numbers usually chosen between $[0,1]$, $c_1$ is a positive constant called as coefficient of the self-recognition component, $c_2$ is a positive constant called as coefficient of the social component (Xin-She Yang, 2010, Dinesh K. Gupta 2015).

6-Results and Performance Analysis

In order to analyze the performance of our proposed de-noising method, MATLAB 2011 program have been used to implement the system shown in Fig.(3). In our algorithm three different signals (sine, rectangular and triangle) are used to test the proposed model, each of which have $N=32000$ symbol length with different frequency range as shown in Figure(4). These signals are contaminated with Gaussian noise with $SNR=10,15,20,$ and $25$ to get a noisy signal from each one.

![Figure (4): The tested signals in simulation.](image-url)

6-1 Kurtosis Statistics of Residual Noise
In this section, a simple level Haar DWT has been used to decompose each noisysignal into approximation and detail coefficients each of which with 1600 samples. The kurtosis function evaluated for detail coefficients after thresholding them by the inverse soft threshold function (Eq. (11)). It is noticed from Figure (5) for each tested signal with different SNR that there is a single value for threshold called optimum threshold \( T_{\text{opt}} \) at which the kurtosis measuring function of residual noise \( R \) be maximum.

**Figure (5):** Kurtosis measuring of residual noise at different SNR levels for (a) Sine wave. (b) Rectangular wave. (c) Triangle wave.

To validate our proposed method, we compare the optimum threshold value that obtained from kurtosis measuring with the well-known one, the BayesShrink threshold (Eq. (9)), the results in Table (1) shows a full agreement between the two threshold values.

<table>
<thead>
<tr>
<th>SNR dB</th>
<th>Sine wave</th>
<th>Rectangle wave</th>
<th>Triangle wave</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( T_{\text{opt}} )</td>
<td>( T_{\text{Bayes}} )</td>
<td>( T_{\text{opt}} )</td>
</tr>
<tr>
<td>10</td>
<td>1.08</td>
<td>1.0934</td>
<td>1.53</td>
</tr>
<tr>
<td>15</td>
<td>0.6</td>
<td>0.6148</td>
<td>0.84</td>
</tr>
<tr>
<td>20</td>
<td>0.33</td>
<td>0.3457</td>
<td>0.48</td>
</tr>
<tr>
<td>25</td>
<td>0.18</td>
<td>0.1944</td>
<td>0.27</td>
</tr>
</tbody>
</table>
6-2 Optimum Threshold value Using PSO Algorithm

PSO algorithm is an effective tool to find these optimum threshold values \( T_{opt} \) in Table (1). For this purpose, the number of particles in the swarm has been considered equal to 10, each particle represents a possible threshold value. Maximum number of iteration assumes to be 100. It should be noted here that the number of iterations that needed to locate maximum kurtosis value depends on time needed for particlesto converge into optimum position.

6-2-1 PSO Algorithm for One Level DWT

Although one level decomposition is rarely used in practical applications, it is used here to check the accuracy of our proposed algorithm. A one level decomposition of a sine wave noisy signal with SNR=10dB has been considered here as an example case. The signal decomposed into approximation and detail coefficients each of which with 16000 samples as shown in Figure (6).

After applying the proposed algorithm, the optimum threshold value was \( T_{PSO} = 1.0602 \) with maximum kurtosis = -1.2051 for the residual noise. The convergence behavior of PSO algorithm (swarm position at each iteration) is shown in Figure (7). Finally \( T_{PSO} \) value used to threshold the detail coefficients and the resultant is reconstructed with approximation coefficients using IDWT to obtain de-noised signal as shown in Figure (8).

![Figure(6): One level DWT decomposition for noisy sine wave at SNR=10dB.](image)

![Figure (7): Convergence behavior of PSO in case of one level DWT decomposition for noisy sine wave at SNR=10dB.](image)
Figure(8): De-noised signal using proposed algorithm in case of one level DWT with noisy sine wave at $SNR=10\,db$.

It's obvious that the de-nosing performance is so awful due to using only one level DWT. The same procedure used for the three signals at four different $SNR$ level to obtain 12 results recorded in Table (2).

Table (2): Threshold values and number of iteration ($n_{iter}$) for PSO algorithm when one level DWT is used.

<table>
<thead>
<tr>
<th>SNR dB</th>
<th>Sine wave $T_{PSO}$</th>
<th>$n_{iter}$</th>
<th>Rectangle wave $T_{PSO}$</th>
<th>$n_{iter}$</th>
<th>Triangle wave $T_{PSO}$</th>
<th>$n_{iter}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.0602</td>
<td>22</td>
<td>1.538</td>
<td>27</td>
<td>0.91216</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>0.59661</td>
<td>23</td>
<td>0.82026</td>
<td>24</td>
<td>0.51365</td>
<td>23</td>
</tr>
<tr>
<td>20</td>
<td>0.33285</td>
<td>24</td>
<td>0.48165</td>
<td>23</td>
<td>0.29531</td>
<td>28</td>
</tr>
<tr>
<td>25</td>
<td>0.19106</td>
<td>28</td>
<td>0.27301</td>
<td>22</td>
<td>0.16519</td>
<td>29</td>
</tr>
</tbody>
</table>

Another validation for the proposed model has been done by comparing $T_{PSO}$ in Table(2) with the corresponding values ($T_{Bayes}$) in Table (1). It's obvious that the PSO algorithm provides a strong agreement results with acceptable computational complexity for all signal types at any SNR level. For that reason our algorithm used such a method that not depending on any statistical properties of the signal to extract the desired threshold value.

6-2-2 PSO Algorithm for Multilevel DWT

In this section DWT with five levels decomposition have been used. The detail coefficients of five sub bands have been chosen (Figure (9)) for a sine wave noisy signal with SNR=10dB as an example case from the three cases for each tested signal.
Figure (9): Five level decomposition for sine wave signal with SNR=10.

After applying the proposed algorithm, the obtained values of maximum kurtosis, threshold and number of iteration for each detail subbands are:

<table>
<thead>
<tr>
<th>Detail coefficient</th>
<th>( n_{\text{iter}} )</th>
<th>( \text{Kurt}_{\text{max}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_{D1} )</td>
<td>23</td>
<td>-1.1784</td>
</tr>
<tr>
<td>( c_{D2} )</td>
<td>22</td>
<td>-1.1971</td>
</tr>
<tr>
<td>( c_{D3} )</td>
<td>24</td>
<td>-1.1794</td>
</tr>
<tr>
<td>( c_{D4} )</td>
<td>21</td>
<td>-1.1665</td>
</tr>
<tr>
<td>( c_{D5} )</td>
<td>21</td>
<td>-1.2034</td>
</tr>
</tbody>
</table>

The convergence behavior of PSO algorithm (swarm position at each iteration) for each detail subbands is shown in Figure (7).
Finally threshold values of $T_{PSO}$ are used to threshold the detail coefficients and the resultant is reconstructed with approximation coefficients using IDWT to obtain de-noised signal as shown in Figure(11).

It's obvious that the five levels DWT provides a powerful performance as compared with one level DWT. For more illustration Table (3) contain mean square error (MSE) of these two cases.

Table (3): MSE of proposed algorithm in case of one and five levels DWT

<table>
<thead>
<tr>
<th>SNR</th>
<th>Sine wave</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Triangle wave</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$MSE_{L5}$</td>
<td>$MSE_{L1}$</td>
<td>$MSE_{L5}$</td>
<td>$MSE_{L1}$</td>
<td>$MSE_{L5}$</td>
<td>$MSE_{L1}$</td>
<td>$MSE_{L5}$</td>
<td>$MSE_{L1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.023984</td>
<td>0.05364</td>
<td>0.016733</td>
<td>0.037642</td>
<td>0.027161</td>
<td>0.063036</td>
<td>0.016733</td>
<td>0.037642</td>
<td>0.027161</td>
<td>0.063036</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.013636</td>
<td>0.030792</td>
<td>0.0094562</td>
<td>0.02206</td>
<td>0.015628</td>
<td>0.036331</td>
<td>0.015628</td>
<td>0.020508</td>
<td>0.015628</td>
<td>0.036331</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.0082766</td>
<td>0.017496</td>
<td>0.0052212</td>
<td>0.012292</td>
<td>0.012292</td>
<td>0.020508</td>
<td>0.0069232</td>
<td>0.011278</td>
<td>0.0069915</td>
<td>0.011278</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.0054104</td>
<td>0.00983</td>
<td>0.0028875</td>
<td>0.0069232</td>
<td>0.0069232</td>
<td>0.011278</td>
<td>0.0069915</td>
<td>0.011278</td>
<td>0.0069915</td>
<td>0.011278</td>
<td></td>
</tr>
</tbody>
</table>
7-Conclusions

This paper proposed a PSO based multilevel adaptivethresholding technique for signal de-noising. Through our work we noticed the following: In Figure (5) it's appeared that: there is a single value for threshold (optimum threshold) which maximizes the kurtosis measuring of residual noise for any signal at any SNR level. This point led us to ensure that our proposed algorithm can be used for any signal without any prior knowledge for its original statistical properties. In section 6-2 PSO algorithm used for searching optimum threshold value. By comparing threshold values that obtained by PSO algorithm in Table (2) with the corresponding values in Table (1) it's obvious that PSO algorithm provides an exultant result with acceptable computational complexity for all signals type at any SNR level. The second important feature for the PSO algorithm that strongly noticed here is the number of iteration needed for the algorithm to convergence to the optimum value is always about (20 ~30) iteration regardless the signal type and SNR level and decomposition level, see Table (2) and Figure (7) and Figure (10). This point can be used to reduce processing time by evaluation only odd value of threshold ($T_1, T_3$,...) and even values ($T_2, T_4,$,...) can be calculated intuitively.

Finally the clear conclusion that the five level DWT provides a more powerful performance than one level decomposition as given in Table (3).

References


Xin-She Yang , 2010, "Engineering Optimization", John Wiley &Sons INC