Wavelet denoising as signal feature-dependent kernel convolution

M R Md Yusof *, A K Ariffin and M Ismail
Department of Physics, Faculty of Science and Mathematics (FSMT), Universiti Pendidikan Sultan Idris (UPSI), Tanjong Malim. Perak. Malaysia

*rozni.yusof@fsmt.upsi.edu.my

Abstract. This work presents a view of wavelet threshold denoising as a convolution process between a noisy signal and a low-pass kernel whose width varies depending on the features of the signal. Wavelet denoising has been widely used for many purposes across a wide range of disciplines, such as in Medical Imaging. However, the procedures are usually presented in a specialized mathematical notation that can be daunting to many researchers, especially those who are not specializing in mathematics. This work seeks to present wavelet denoising in a form more accessible to researchers in other disciplines such as Geographical Information Systems, image processing, and remote sensing.

1. Introduction
Wavelet techniques have been used for denoising of signals in many areas of research. It has been used for denoising medical images [1], reflectance spectra [2-4], astronomical signals [5] electrocardiograms (ECG) signals [6] and also GIS and remote sensing [7]. Other applications include denoising ultra high frequency signals in monitoring partial discharge signals [8] and denoising of hyperspectral image cubes [7]. The wavelet transform denoising technique is said to be the most optimal signal denoising. It provides removal of noise while preserving sharp features of the signal [9,10].

In terms of algorithms, the Stationary Wavelet Transform (SWT) was shown to be superior to the Mallat Algorithm in denoising [11]. More recent developments include the “Lifting Algorithm” [12].

Wavelet denoising is often presented in specialized mathematical notation which forms a steep learning curve for most researchers. Other research present the process by simply stating the wavelet algorithm used, how wavelets are implemented digitally, and the threshold used, without going into depth.

This work presents wavelet denoising as convolution of a signal with an adaptive low-pass kernel. The advantage of this approach is that it is more intuitive and only requires researchers to understand the effect of convolution of a signal with a kernel. First the denoising process is described using wavelets that were dilated continuously. Next the effect of using SWT to implement wavelet denoising is shown, in terms of how it affects the kernel to be used for denoising.

2. Theory

2.1 Wavelet Decomposition
The continuous wavelet transform is given as
\[ W_a(b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi \left( \frac{t-b}{a} \right) dt \]  

The result is a set of wavelet coefficients \( W_k(t) \) for level \( b \) [13]. The wavelet kernel is then dilated to a different scale of \( a \) and the signal is convolved with the dilated kernel to generate another set of wavelet coefficients, for as many levels as desired.

The wavelet transform can be described as a convolution between a signal \( f(t) \) and a wavelet function \( \psi(t) \), such that:

\[ W_a(t) = f(t) * \psi_a(t) \]  

where \( a \) denotes the scale of the wavelet. The result is a set wavelet coefficients \( W_a(t) \). The wavelet function \( \psi_a(t) \) is dilated such that

\[ \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t-b}{a} \right) \]  

Discrete wavelet transform algorithms such as the Mallat Algorithms and the Stationary Wavelet Transform are implemented using wavelet filter banks on dyadic scales [9,11,13]. The algorithms employ wavelet and approximation kernels, denoted \( \psi \) and \( \phi \), which are dilated by a factor of two at each scale. The signal \( f(t) \) is first convolved with the kernels \( \psi \) and \( \phi \) to generate wavelet detail coefficients \( d_1 \) and wavelet approximation coefficients \( a_1 \). The functions \( \psi \) and \( \phi \) are dilated by a factor of two, and then applied to the approximation coefficients \( a_1 \), to produce the detail and approximation coefficients for the second level \( d_2 \) and \( a_2 \). The process is repeated for as several levels. The maximum number of levels of decomposition is given as \( \log_2(\text{length}) \), where \( \text{length} \) denotes the number of data points in the signal. The discrete wavelet transform is usually performed on data whose length is a power of two.

Figure 1 shows a flow chart for three levels of decomposition.

![Figure 1. Flow chart for 3 levels of wavelet decomposition](image)

The dilated kernels for the \( j^{th} \) level are denoted:

\[ \psi_j = \frac{1}{\sqrt{2^j}} \psi \left( \frac{t-2^j k}{2^j} \right) \]

\[ \phi_j = \frac{1}{\sqrt{2^j}} \phi \left( \frac{t-2^j k}{2^j} \right) \]  

The kernels are generated from the parent wavelet kernels \( \phi \) and \( \psi \).
The actual method of dilating the wavelet kernels depends on the algorithm being used. The *a-trous* algorithm, better known as the Stationary Wavelet Transform (SWT) dilates the wavelets by inserting \((2^j - 1)\) zeros between each point in the wavelet high pass and low pass kernels, where \(j\) is the level of decomposition. We explore wavelet denoising using the continuous dyadic expressions for the high pass and low pass kernels as given in the equations above [14].

From Figure 1 we can see that the generation of detail and approximation coefficients at each level involve convolving the signal with a sequence of high pass and low pass filters. For instance, for the first three levels of decomposition:

\[
d_1 = \psi \ast f \\
d_2 = \psi_2 \ast \phi_1 \ast f \\
d_3 = \psi_3 \ast \phi_2 \ast \phi_1 \ast f \\
a_3 = \phi_3 \ast \phi_2 \ast \phi_1 \ast f
\]

The signal \(f\) can be reconstructed using the wavelet reconstruction kernels \(\phi'\) and \(\psi'\), convolving the detail and approximation coefficients with them and adding them in the appropriate sequence:

\[
f = \psi'_1 \ast d_1 + \phi_1 \ast \psi'_2 \ast d_2 + \phi'_1 \ast \phi'_2 \ast \psi'_3 \ast d_3 + \phi'_1 \ast \phi'_2 \ast \phi'_3 \ast a_3
\]

The reconstruction kernels are simply the time-reversed decomposition filters. This can be generalized, for \(j=1\) to \(N\) levels of decomposition borrowing the symbol \(\prod\) normally used for serial multiplication to denote serial convolution, so that \(\prod_{i=1}^{N} \phi_i = \phi_1 \ast \phi_2 \ast \ldots \ast \phi_N\):

\[
d_1 = \psi_1 \ast f \\
d_j = \psi_j \ast \prod_{i=2}^{j-1} (\phi_i \ast) f \\
a_N = \prod_{j=1}^{N} (\phi_j \ast) f
\]

while for reconstruction:

\[
f = \psi'_1 \ast d_1 + \sum_{j=2}^{N} \psi'_j \ast \prod_{i=1}^{j-1} (\phi_i \ast) d_j + \prod_{j=1}^{N} (\phi_j \ast) a_N
\]

### 2.2. Denoising using Wavelet Thresholding

Wavelet thresholding is one method of using the wavelet transform to remove noise from a signal. The process involves modifying the wavelet detail coefficients by applying a threshold rule, such that

\[
\hat{d}_j(t) = \begin{cases} 
  d_j(t) & \text{if } |d_j(t)| \geq T \\
  0 & \text{if } |d_j(t)| < T 
\end{cases}
\]

where \(T\) is a threshold value. This can also be written

\[
\hat{d}_j(t) = \delta_j(t) d_j(t)
\]

where

\[
\delta_j = \begin{cases} 
  1 & \text{if } |d_j| \geq T \\
  0 & \text{if } |d_j| < T 
\end{cases}
\]

The threshold value is computed by various methods known by various names, and are implemented in software toolboxes in Matlab. Commonly implemented thresholding rules include Fixed Form Threshold [9], SURE and Hybrid SURE [10].
We give the Fixed Form Threshold as:

\[ T = \sigma \sqrt{2 \log(\text{length})} \]  

(11)

where \( \sigma \) is the standard deviation of the noise and \( \text{length} \) is the number of points in the signal we are denoising. The standard deviation is, in practice, estimated by way of the median absolute deviation (MAD) of the first level detail coefficients. This is the simplest of the threshold rules [15].

The denoised signal is thus the result of reconstructing the signal using the modified detail coefficients:

\[
\hat{f} = \psi_1' \hat{a}_1 + \sum_{j=2}^{N} \psi_j' * \prod_{i=1}^{j-1} (\phi_i *) \hat{d}_j + \prod_{j=1}^{N} (\phi_j *) a_N
\]  

(12)

When combined with Error! Reference source not found.:

\[
\hat{f} = \delta_1 \psi_1' * \psi_1 * f + \sum_{j=1}^{N} [\delta_j (\psi_j' * \psi_j) * \prod_{i=1}^{j-1} (\phi_i * \phi_i)] f + \prod_{j=1}^{N} (\phi_j * \phi_j) f
\]  

(13)

We may think of \( A \) as a convolution kernel that we use to process and denoise the signal \( f(t) \). The final term \( \prod_{j=1}^{N} (\phi_j * \phi_j) \) results in a low pass filter that actually performs the smoothing.

When the equations given in Error! Reference source not found. and Error! Reference source not found. are applied to continuous wavelets, the results are shown in Figure 2, Figure 3 and Figure 4. Figure 2 shows a low pass filter kernel generated by \( \prod_{j=1}^{3} (\phi_j' * \phi_j *) \) where \( \phi_j \) and \( \phi_j' \) are the decomposition and reconstruction approximation filters of the Daubechies db4 wavelet. This is expanded as \((\phi_1' * \phi_1) * (\phi_2' * \phi_2) * (\phi_3' * \phi_3)\):

![Figure 2. Low-pass kernel for 3 levels of decomposition](image)

One of the problems of denoising using a low-pass filter such as boxcar, triangle or Gaussian kernel is that it distorts sharp features in the signal, by indiscriminately smoothing out every feature. Wider kernels tend to excel at removing noise but will also remove sharp features.

However, in wavelet denoising, the first two terms of \( A \) mitigate this by altering the low-pass filter at sharp features in the signal, making the low-pass filter kernel narrower and preserving the sharp features. The two terms will depend on the wavelet detail coefficients \( \delta_j \) being nonzero. For smooth
regions of the signal, $\delta_j(t)$ will mostly be zero and the two terms do not modify the low pass filter in Figure 2.

However, at sharp regions of the signal, $\delta_j(t)$ will be nonzero and the first two terms modify the low-pass kernel produced in the third term. At the extreme, if all the terms are added, the low pass filter is modified into a kernel that approaches a unit impulse function, which simply reproduces the signal it is convolved with without smoothing or other alteration. Thus, we may think of wavelet denoising as a deconvolution process with a lowpass kernel that adapts its width depending on signal feature.

Figure 3 shows the result of $\psi_1 * \psi_1 + \sum_{j=1}^3 (\psi_j * \psi_j) * \prod_{i=1}^{j-1} (\phi_i' * \phi_i)$ for $1 \leq j \leq 3$, $j \in \mathbb{Z}$, if $\delta_j = 1$. From top to bottom, it shows the resulting function for the expansions $\psi_1 * \psi_1$, followed by $\psi_1' * \psi_2 + (\psi_2' * \psi_2) * (\phi_1' * \phi_1)$ and finally $\psi_1 * \psi_1 + (\psi_2' * \psi_2) * (\phi_1' * \phi_1) + (\psi_3' * \psi_3) * (\phi_1' * \phi_1) * (\phi_2' * \phi_2)$.

Figure 4 shows the result of progressively adding these kernels to the original low-pass kernel in Figure 2. From top to bottom, the Figure shows the result of adding the level 3 high-pass modifier, followed by the level 2 and level 1 high pass kernels.

The addition of the high-pass terms cause the low-pass filter to become narrower, approaching the unit impulse function, which allows the high-frequency components to pass and thus preserve the sharp features in the signal.
2.3. Effect of Implementation in Stationary Wavelet Transform

The Stationary Wavelet Transform Algorithm is an implementation of the wavelet transform at dyadic scales, where the kernels $\psi, \psi', \phi$ and $\phi'$ are dilated by inserting $2^j - 1$ zeros between each term of the wavelet kernel, where $j$ is the level of decomposition. The $k^{th}$ term of the decomposition kernels $\phi$ and $\psi$ at the $i^{th}$ level is such that:

$$\psi_j[2^{j-1}k] = \psi_0[k]$$

$$\phi_j[2^{j-1}k] = \phi_0[k]$$  \(15\)

where $\psi_0$ and $\phi_0$ are the original undilated kernels. The reconstruction kernels are the time reversals of the above [13]. The algorithm used to implement wavelet decomposition has an effect on the shape of the kernel.

Figure 5 and Figure 6 show the contrast when Error! Reference source not found. and Error! Reference source not found. are applied to the Daubechies $db1$ wavelet, first to wavelet dilated in a
continuous way and then for the wavelet dilated by inserting zeros as in SWT. The \textit{db1} wavelet, also known as \textit{Haar}, was chosen because the effect of the SWT algorithm is most apparent in this wavelet.

Figure 5 shows the result for \textit{db1} wavelet dilated continuously. The topmost figure low-pass kernel generated for 3 levels of decomposition. The middle figure shows the high-pass modifier whose existence depends on the features of the signal, while the lowest figure shows the resultant kernel from adding the two kernels. The result for the low-pass kernels shows almost an ideal low-pass kernel shape, very closely approximating the Gaussian kernel low-pass filter.

![Figure 5](image)

**Figure 5.** Low-pass kernel, modifier and modified low-pass kernel generated for \textit{db1} wavelet using continuous wavelet function

In contrast, Figure 6 shows the result for the \textit{db1} wavelet where the kernels are dilated by inserting zeros in accordance with the SWT algorithm. The topmost figure in Figure 6 shows the low-pass kernel generated for 3 levels of decomposition. Instead of a smooth nearly ideal low-pass kernel, the result is a triangular low-pass kernel. The middle figure shows the modifying kernel, while the lowest figure shows the result of adding the first and second kernels, which is a narrower triangle kernel.
Figure 6. Low-pass kernel, kernel modifier and resulting kernel generated using db1 wavelet, using Stationary Wavelet Transform algorithm

The triangle kernel is not quite as good as a Gaussian kernel in terms of denoising. This can be seen by performing Fourier analysis of the two kernels, which will show the triangle kernel having small lobes in the frequency domain. Using higher order wavelets such as db2 or db4 goes some way to mitigate this problem.

3. Conclusion
In this work, wavelet denoising has been described as a convolution between a signal and a low-pass kernel that varies its width depending on the features of the signal, which allows sharp features in the signal to be preserved while filtering noise. This characteristic of wavelet denoising is what enables it to be a very effective noise removal technique. The effect of using SWT to implement wavelet denoising has also been demonstrated for the db1 or Haar wavelet.

Further work that needs to be done include the effect of non-constant noise variance, which describes photon detection systems such as CCD [16], and the effect of more recently developed wavelet algorithms such as the Lifting Algorithm on the wavelet kernel.

Acknowledgement
This research was made possible by the University Research Grant (GPU) 2016-0189-102-01, through the Research Management and Innovation Committee (RMIC) of Universiti Pendidikan Sultan Idris (UPSI).

References
[1] Nowak, R. D., & Baraniuk, R. G. (oct, 1999). Wavelet Domain Filtering for Photon Imaging Systems. In A. Aldroubi, A. F. Laine, & M. A. Unser (Ed.), Wavelet Applications in Signal and Image Processing V. 5. SPIE-Intl Soc Optical Eng. doi:10.1117/12.279679

[2] Shafri, H. Z., & Mather, P. M. (jul, 2005). Wavelet Shrinkage in Noise Removal of Hyperspectral Remote Sensing Data. American Journal of Applied Sciences, 2, 1169-1173. doi:10.3844/ajassp.2005.1169.1173
[3] Schmidt, K. S., & Skidmore, A. K. (mar, 2004). Smooth Vegetation Spectra with Wavelets. *International Journal of Remote Sensing*, Vol. 25, No.6, pp.1167-1184, Taylor and Francis, 25, 1167-1184. doi:10.1080/014316031000115085

[4] Shafri, H. Z., & Mather, P. M. (2002). Wavelet-based de-noising for derivative spectra analysis. IOS Press. Retrieved from https://www.amazon.com/Geoinformation-European-wide-Integration-Proceedings-Symposium/dp/9077017712%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D9077017712

[5] Kolaczyk, E. (1996). Estimation of Intensities of Burst-Like Poisson Processes Using Haar Wavelets. Tech. rep., Biometrika. doi:10.1.1.57.2213

[6] Rafiee, J., Rafiee, M. A., Prause, N., & Schoen, M. P. (may, 2011). Wavelet basis functions in biomedical signal processing. *Expert Systems with Applications*, 38, 6190-6201. doi:10.1016/j.eswa.2010.11.050

[7] Sulochana, S., Vidhya, R., Vijayasekaran, D., & Maharaj, K. (8, 2016). Denoising and Dimensionality Reduction of Hyperspectral Images using Framelet Transform with Different Shrinkage Functions. *Indian Journal of Geo-Marine Sciences*, 45, 978-986. Retrieved from http://nopr.niscair.res.in/handle/123456789/35182

[8] Yang, L., Judd, M. D., & Bennoch, C. J. (10, 2004). Denoising UHF signal for PD detection in transformers based on wavelet technique. *The 17th Annual Meeting of the [IEEE] Lasers and Electro-Optics Society, 2004. [LEOS] 2004. (pp. 166-169). Institute of Electrical and Electronics Engineers (IEEE). doi:10.1109/CEIDP.2004.1364215

[9] Donoho, D. L. (may, 1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41, 613-627. doi:10.1109/18.382009

[10] Donoho, D. L., & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 425-455. doi:10.1093/biomet/81.3.425

[11] Coifman, R. R., & Donoho, D. L. (1995). Translation-Invariant De-Noising. 125-150. doi:10.1007/978-1-4612-2544-7_9

[12] Ebadi, L., Shafri, H. Z., Mansor, S. B., & Ashurov, R. (mar, 2013). A review of applying second-generation wavelets for noise removal from remote sensing data. *Environmental Earth Science, Springer Berlin Heidelberg*, 70, 2679-2690. doi:10.1007/s12665-013-2325-z

[13] Mallat, S. (1999). *A Wavelet Tour of Signal Processing, Second Edition (Wavelet Analysis & Its Applications)*. Academic Press.

[14] Pizurica, A. (2002). *Image Denoising Using Wavelets and Spatial Context Modelling*. Ph.D. dissertation, University of Ghent.

[15] Singh, B. N., & Tiwari, A. K. (may, 2006). Optimal selection of wavelet basis function applied to ECG signal denoising. *Digital Signal Processing, Vol. 16, Issue 3, pp 275-287, Academic Press, 16, 275-287.* doi:10.1016/j.dsp.2005.12.003

[16] Starck, J. L., & Murtagh, F. (mar, 2001). Astronomical image and signal processing: looking at noise, information and scale. *IEEE Signal Processing Magazine, March, 18*, 30-40. doi:10.1109/79.916319

[17] Yusof, M. R. (may, 2009). Determination of optimal wavelet denoising parameters for red edge feature extraction from hyperspectral data. *Journal of Applied Remote Sensing*, 3, 033533. doi:10.1117/1.3155804