We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



4

Wavelet Denoising

Guomin Luo and Daming Zhang
Nanyang Technological University
Singapore

1. Introduction

Removing noise from signals is possible only if some prior information is available. The information is employed by different estimators to recover the signal and reduce noise. Most noises in one-dimensional transient signal follow Gaussian distribution. The Bayes estimator minimizes the expected risk to get the optimal estimation. The minimax estimator uses a simple model for estimation. They are the most popular estimators in noise estimation.

No matter which estimator is used, the risk should be as small as possible. Donoho and Johnstone have made a breakthrough by proving that thresholding estimator has a small risk which is close to the lower bound. Thereafter, threshold estimation was studied extensively and has been improved by more and more researchers. Besides the universal threshold, some other thresholds, for example SURE threshold and minimax threshold, are also widely applied.

In wavelet denoising, the thresholding algorithm is usually used in orthogonal decompositions: multi-resolution analysis and wavelet packet transform. Wavelet thresholding faces some questions in its application, for example, the selection of hard or soft threshold, fixed or level-dependent threshold. Proper selection of those items helps generating a better estimation.

Besides the influence of thresholding, the influence of wavelets is also an important factor. In most applications, the wavelet transform uses a few non-zero coefficients to describe a signal or function. Producing only a few non-zero coefficients is crucial in noise removal and reducing computing complexity. Choosing a wavelet with optimum design to produce more wavelet coefficients close to zero is crucial in some applications.

2. Noise estimation

The output acquired by sensing devices, for example transformer and sensor, is a measurement of analogue input signal $\bar{f}(x)$. The output can be modelled as in (1). The output $X[n]$ is composed by a filtered $\bar{f}(x)$ with the sensor responses $\bar{\phi}(x)$ and an added noise $W[n]$. The noise W contains various types of interferences, for instance, the radio frequency interferences from communication systems. It also includes the noises induced by measurement devices, such as electronic noises from oscilloscope and transmission errors. In most cases, the noise W is modelled by a random vector that often follows Gaussian distribution.

$$X[n] = < \overline{f}, \overline{\phi} > + W[n] \tag{1}$$

If the analogue-to-digital data acquisition is stable, the discrete signal can be denoted by $f[n] = < \overline{f}, \overline{\phi} >$. The analogue approximation of $\overline{f}(x)$ can be recovered from $f[n]$. The noisy output in (1) is rewritten as

$$X[n] = f[n] + W[n] \tag{2}$$

The estimation of $f[n]$ calculated from (2) is denoted by $\tilde{F} = DX[n]$, where D is the decision operator. It is designed to minimize the error $f - \tilde{F}$. For one-dimension signal, the mean-square distance is often employed to measure the error $f - \tilde{F}$. The mean-square distance is not a perfect model but it is simple and sufficiently precise in most cases (Mallat, 2009d). The risk of the estimation is calculated by (3):

$$r(D, f) = E\{\| f - \tilde{F} \|^2\} \tag{3}$$

The decision operator D is optimized with the prior information available on the signal (Mallat, 2009d). Two estimation methods: Bayes estimation and minimax estimation are the most commonly used ones. The Bayes estimator minimizes the risk to get optimal estimation. But it is difficult to obtain enough information to model prior probability distribution. The minimax estimator uses simple model. But the risk cannot be calculated. The section 2.1 and section 2.2 introduce the fundamental of Bayes estimator and minimax estimator.

2.1 Bayes estimation

In Bayes estimation, the unknown signal f which is denoted by a random vector F is supposed to have a probability distribution α which is also called prior distribution. The noisy output in (2) can be rewritten as

$$X[n] = F[n] + W[n] \text{ for } 0 \le n < N \tag{4}$$

The noise W is supposed to be independent with F for all n. The distribution of measurement X is the joint distribution of F and W. It is called posterior distribution. Thus, $\tilde{F} = DX$ is an estimator of F from measurement X. Then the risk is the same as in (3). The Bayes risk of $\tilde{F}$ with respect to the prior probability distribution α of the signal is:

$$r(D, \alpha) = E_\alpha\{r(D, F)\} = E\{\| F - \tilde{F} \|^2\} = \sum_{0}^{N-1} E\{ | F[n] - \tilde{F}[n] |^2\} . \tag{5}$$

The estimator $\tilde{F}$ is said to be a Bayes estimator if it minimizes the Bayes risk among all estimators. Equivalently, the estimator which minimizes the posterior expected loss $E_\alpha\{r(D, F) | X\}$ for each X also minimizes the Bayes risk and therefore is a Bayes estimator (Lehmann & Casella, 1998).

The risk function is determined by choosing the way to measure the distance between the estimator $\tilde{F}$ and the unknown signal F. In most applications, the mean square error is adopted because of its simplicity. But some alternative estimators are also used such as

linear estimation. In this chapter, most estimations use mean square error to measure estimation risk.

2.2 Minimax estimation

It is possible that we have some prior information for a signal, but it is rare to know the probability distribution of complex signals. For example, there is not an appropriate model for the stochastic transient signals in power system or the sound signals from nature environment. In this case, we have to find a "good" estimator whose maximal risk is minimal among all estimators. The prior information forms a signal set Θ. But this set does not specify the probability distribution of signals in Θ. The more prior information, the smaller the set Θ (Mallat, 2009d).

For the signal $f \in \Theta$, the noisy output is as shown in (2). The risk of estimation $\tilde{F} = DX$ is $r(D, f) = E\{\| DX - f \|^2\}$. Since the probability distribution of signal in set Θ is unknown, the precise risk cannot be calculated. Only a possible range is calculated. The maximum risk of this range is (Donoho & Johnstone, 1998):

$$r(D, \Theta) = \sup_{f \in \Theta} E\{\| DX - f \|^2\} \tag{6}$$

In minimax estimation, the minimax risk is the lower bound of risk in (6) with all possible, no matter linear or nonlinear, operators D:

$$r_n(D, \Theta) = \inf_{D \in On} r(D, \Theta) \tag{7}$$

Here, O_n denotes the set of all operators.

3. Threshold estimation in bases

Threshold is the estimated noise level. The values larger than threshold are regarded as signal, and the smaller ones are regarded as noises. When the noisy output is decomposed in a chosen base, the estimator of noises can also be applied because the white noises remain as white noises in all bases. It is proved in section 3.1. Two thresholding functions: hard thresholding and soft thresholding are introduced in section 3.2.

3.1 Estimation in orthogonal basis

When the noisy output is decomposed in an orthogonal basis $B = \{g_m\}_{0 \le m < N}$, the components in (2) is rewritten as $X_B[m] = < X, g_m >$, $f_B[m] = < f, g_m >$, and $W_B[m] = < W, g_m >$. The sum of them gives

$$X_B[m] = f_B[m] + W_B[m]. \tag{8}$$

If the noise W is a zero-mean white noise with variance σ^2, then $E\{W[n]W[k]\} = \sigma^2 \delta[n-k]$. Thus the noise coefficients $W_B[m] = \sum_{n=0}^{N-1} W[n] g_m^*[n]$ also represent a white noise of variance σ^2. This because,

$$E\{W_B[m]W_B[p]\} = \sum_{n=0}^{N-1}\sum_{k=0}^{N-1} g_m[n]g_p[k]E\{W[n]W[k]\} = \sigma^2 < g_p, g_m >= \sigma^2\delta[p-m]\cdot \quad (9)$$

From the analysis above, one can see that the noise remains as white noise in all bases. It is not influenced by the choice of basis (Mallat, 2009d).

3.2 Thresholding estimation

In the orthogonal basis $B = \{g_m\}_{0\le m<N}$, the estimator of f in $X = f + W$ can be written as:

$$\tilde{F} = DX = \sum_{m=0}^{N-1} a_m(X_B[m])X_B[m]g_m \cdot \quad (10)$$

Here, a_m is the thresholding function. It could be hard thresholding or soft thresholding.

3.2.1 Hard thresholding

A hard thresholding function is shown as follows (Mallat, 2009d):

$$a_m(x) = \begin{cases} 1 & \text{if} \mid x \mid \ge T \\ 0 & \text{if} \mid x \mid < T \end{cases}. \quad (11)$$

By substituting $a_m(x)$ into (10), we can obtain the estimator with hard thresholding function

$$\tilde{F} = \sum_{m=0}^{N-1} \rho_T(X_B[m])g_m \quad \text{with} \quad \rho_T(x) = a_m(x) * x = \begin{cases} x & \text{if} \mid x \mid \ge T \\ 0 & \text{if} \mid x \mid < T \end{cases}. \quad (12)$$

Then the risk of this thresholding is

$$r_{th}(f) = r(D, f) = \sum_{m=0}^{N-1} E\{\mid f_B[m] - \rho_T(X_B[m]) \mid^2\} \cdot \quad (13)$$

3.2.2 Soft thresholding

A soft thresholding function is implemented by (Mallat, 2009d)

$$0 \le a_m(x) = \max(1 - \frac{T}{\mid x \mid}, 0) \le 1 \cdot \quad (14)$$

The resulting estimator $\tilde{F}$ for this case can be written as in (12) with the thresholding function ρ_T replaced by a soft thresholding function as shown in (15).

$$\rho_T(x) = \begin{cases} x - T & \text{if } x \ge T \\ x + T & \text{if } x \le -T \\ 0 & \text{if } \mid x \mid \le T \end{cases} \cdot \quad (15)$$

Reducing the magnitude of coefficients X_B that are greater than threshold usually makes the amplitude of the estimated signal $\tilde{F}$ be smaller than the original F. This is intolerable

for some applications where precise recovery is required such as noise reduction of partial discharge signal, since the pulse magnitude and shape in such applications are needed for further analysis (Zhang et al., 2007). For other cases where precise recovery of signal magnitude is not required, for example, image noise reduction, the soft thresholding is widely used since it can retain the regularity of signal (Donoho, 1995).

The $\rho_T(x)$ of hard thresholding and soft thresholding are portrayed in Fig.1.

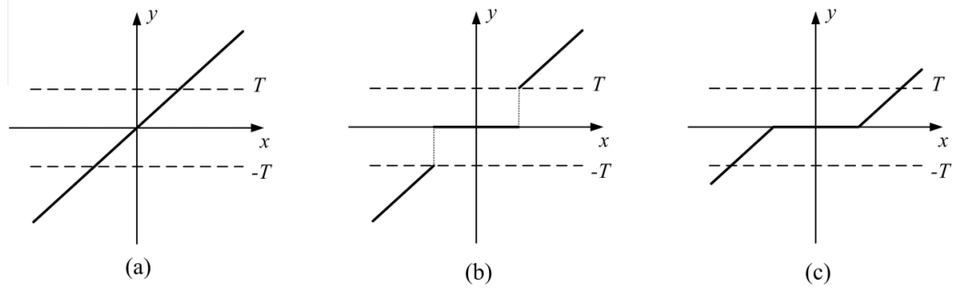


Fig. 1. Thresholding function $\rho_T(x)$, (a) original signal, (b) hard thresholding, (c) soft thresholding

3.3 Threshold estimation and its improvement

As depicted in (13), the risk of thresholding is closely related to the thresholding function ρ_T. The appropriate choice of threshold T is an important factor to reduce the risk of estimation. Several famous thresholds were proposed and proved by different estimation methods.

3.3.1 Universal threshold

Donoho and Johnstone (Donoho & Johnstone, 1994) proposed a universal threshold T. They proved that the risk of thresholding, no matter hard or soft, is small enough to satisfy the requirements of most applications.

If the thresholding function $\rho_T(x) = (x - \lambda sign(x))_{|x|>\lambda}$ is a soft thresholding, for a Gaussian random variable X of mean μ and variance 1, then the estimation risk is

$$r(\lambda, \mu) = E\{|\rho_\lambda(X) - \mu|^2\}. \tag{16}$$

If X has a variance σ^2 and a mean μ, then the following formula can be verified by considering $\tilde{X} = X / \sigma$

$$E\{|\rho_\lambda(X) - \mu|^2\} = \sigma^2 r(\frac{\lambda}{\sigma}, \frac{\mu}{\sigma}). \tag{17}$$

Since the projection $f_B[m]$ in basis $B = \{g_m\}_{0 \le m < N}$ is a constant, the $X_B[m]$ is a Gaussian random variable with mean $f_B[m]$ and variance σ^2. The risk of estimation by soft thresholding with a threshold T is

$$\begin{aligned} r_{th}(f) &= \sigma^2 \sum_{m=0}^{N-1} r(\frac{T}{\sigma}, \frac{f_B[m]}{\sigma}) \\ &\le N\sigma^2 r(\frac{T}{\sigma}, 0) + \sigma^2 \sum_{m=0}^{N-1} \min(\frac{T^2+\sigma^2}{\sigma^2}, \frac{|f_B[m]|^2}{\sigma^2}) \end{aligned} . \tag{18}$$

Donoho and Johnstone proved that for $T = \sigma\sqrt{2\log_e N}$ and $N \ge 4$, the upper bound of the two parts of risk in (18) are (Donoho &Johnstone, 1994)

$$N\sigma^2 r(\frac{T}{\sigma}, 0) \le (2\log_e N + 1)\sigma^2 , \tag{19}$$

and

$$\sigma^2 \sum_{m=0}^{N-1} \min(\frac{T^2+\sigma^2}{\sigma^2}, \frac{|f_B[m]|^2}{\sigma^2}) \le (2\log_e N + 1) \sum_{m=0}^{N-1} \min(\sigma^2, |f_B[m]|^2) . \tag{20}$$

Then, the risk of estimator with threshold $T = \sigma\sqrt{2\log_e N}$ and all $N \ge 4$ is

$$r_{th}(f) \le (2\log_e N + 1)(\sigma^2 + \sum_{m=0}^{N-1} \min(\sigma^2, |f_B[m]|^2)) . \tag{21}$$

Donoho and Johnstone also mentioned in (Donoho &Johnstone, 1994), the universal threshold is optimal in certain cases defined by (Donoho &Johnstone, 1994).

3.3.2 SURE threshold

The thresholding risk is often reduced by decreasing the value of threshold, for instance, choosing a threshold smaller than universal threshold in section 3.3.1. Sure threshold was proposed by Stein (Stein, 1981) to suit this purpose.

As depicted in (Mallat, 2009d), when $|X_B[m]| < T$, the coefficient is set to zero by soft thresholding. Then the risk of estimation equals $|f_B[m]|^2$. Since $E\{|X_B[m]|^2\} = |f_B[m]|^2 + \sigma^2$, the $|f_B[m]|^2$ can be estimated with $|X_B[m]|^2 - \sigma^2$. But if $|X_B[m]| \ge T$, the soft thresholding substracts T from $|X_B[m]|$. Then the risk is the sum of noise energy and the bias introduced by the reduction of the amplitude of $X_B[m]$ by T. The resulting estimator of $r_{th}(f)$ is

$$Sure(X,T) = \sum_{m=0}^{N-1} C(X_B[m]) , \tag{22}$$

with

$$C(u) = \begin{cases} u^2 - \sigma^2 & \text{if } u \le T \\ \sigma^2 + T^2 & \text{if } u > T \end{cases} . \tag{23}$$

The $Sure(X,T)$ is called Stein unbiased risk estimator (SURE) and was proved unbiased by (Donoho & Johnstone, 1995). Consider using this estimator of risk to select a threshold:

$$\tilde{T} = \arg\min_{T} Sure(X,T) \tag{24}$$

Arguing heuristically, one expects that, for large dimension N, a sort of statistical regularity will set in, the Law of Large Numbers will ensure that SURE is close to the true risk and that $\tilde{T}$ will be almost the optimal threshold for the case at hand (Donoho & Johnstone, 1995).

Although the SURE threshold is unbiased, its variance will induce errors when the signal energy is smaller than noise energy. In this case, the universal threshold must be imposed to remove all the noises. Since $E\{\| X \|^2\} = \| f \|^2 + N\sigma^2$, $\| f \|^2$ can be estimated by $\| X \|^2 - N\sigma^2$ and compared with a minimum energy level $\varepsilon_N = \sigma^2 N^{1/2} (\log_e N)^{3/2}$. Then the SURE threshold is (Mallat, 2009d)

$$T = \begin{cases} \sigma\sqrt{2\log_e N} & \text{if } \| X \|^2 - N\sigma^2 \le \varepsilon N \\ \tilde{T} & \text{if } \| X \|^2 - N\sigma^2 > \varepsilon N \end{cases}. \tag{25}$$

3.3.3 Minimax threshold

The inequality in (21) shows that the risk can be represented in the form of a multiplication of a constant $2\log_e N + 1$ and the loss for estimation. However, it is natural and more revealing to look for 'optimal' thresholds λ which yield smallest possible constant Λ in place of $2\log_e N + 1$. Thus, the inequality in (21) can be rewritten as

$$r_{th}(f) \le \Lambda(\sigma^2 + \sum_{m=0}^{N-1} \min(\sigma^2, |f_B[m]|^2)) . \tag{26}$$

The minimax estimation introduced in section 2.2 is a possible method to find the appropriate constant Λ that satisfies $\Lambda \le 2\log_e N + 1$, and the threshold $\lambda \le \sqrt{2\log_e N}$.

Donoho and Jonestone (Donoho & Johnstone, 1994) defined the minimax quantities

$$\Lambda \equiv \inf_{\lambda} \sup_{\mu} \frac{\rho_T(\lambda,\mu)}{N^{-1} + \min(\mu^2, 1)}, \text{ and } T \equiv \text{the largest } \lambda \text{ attaining } \Lambda \text{ above} . \tag{27}$$

They also proved that Λ attains its maximum Λ^0 at $\mu = 0$. Then T is the largest λ attaining Λ^0. Since $\rho(\lambda,\infty)$ is strictly increasing in λ and $\rho(\lambda,0)$ is strictly decreasing in λ, so that the solution of (27) is

$$(N+1)\rho_T(\lambda,0) = \rho_T(\lambda,\infty) . \tag{28}$$

Then with this solution, the minimax threshold T is

$$T \le \sqrt{2\log_e N}, \quad T^2 = 2\log_e(N+1) - 4\log_e(\log_e(N+1)) - \log_e 2\pi + o(1) \quad (N \to \infty) . \tag{29}$$

Usually, for the same N, the risk of universal threshold is larger than SURE threshold and minimax threshold. All the three thresholds mentioned in section 3.3.1 to section 3.3.3 are

applied to denoise the same noisy data and are evaluated by signal-to-noise ratio (SNR), which is measured in decibels:

$$SNR_{dB} = 10 * \log_{10}(\frac{E\{\| F \|^2\}}{E\{\| F - \tilde{F} \|^2\}}), \tag{30}$$

where F is the original data without noise and $\tilde{F}$ is the estimation of F.

Fig.2 shows the estimation of a synthesized signal with different thresholds. The noisy data is decomposed in a biorthogonal basis. Since hard thresholding is adopted, setting a wavelet coefficient to zero will induce oscillations near discontinuities in estimation. The estimation with universal threshold in Fig.2(c) shows small oscillations at the smooth parts.

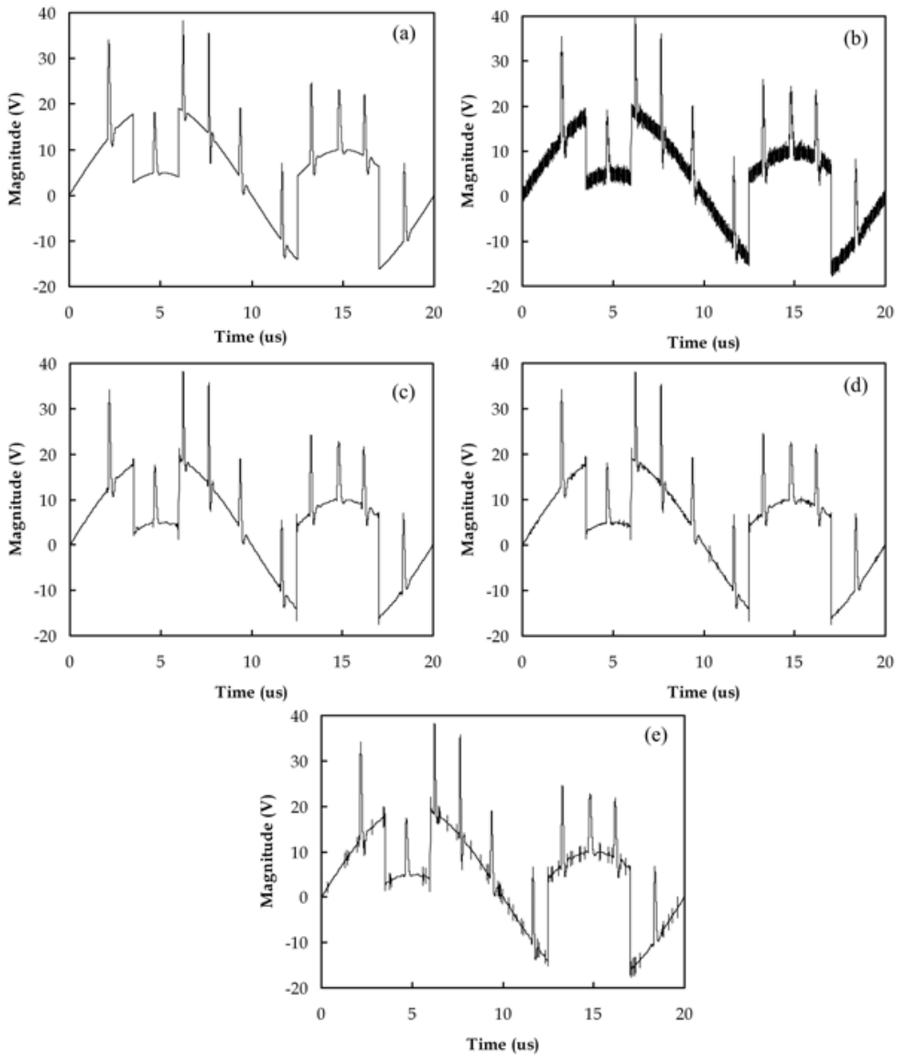


Fig. 2. Estimation with different thresholds. (a) original data, (b) noisy data (SNR=23.59dB), (c) estimation with universal threshold (SNR=31.98dB), (d) estimation with SURE threshold (SNR=34.82dB), (e) estimation with minimax threshold (SNR=33.63dB)

The oscillations result in a smaller SNR (31.98dB). The oscillations are less obvious in estimations in Fig.2(d) and Fig.2(e). But noise with very small magnitude is still found. As mentioned before, universal threshold is usually larger than the other two thresholds. Its risk of estimation $r = E\{\| F - \tilde{F} \|^2\}$ is therefore greater than that of the other two. This can be deduced by values of SNR.

4. Wavelet thresholding

The signals carry a large amount of useful information which is difficult to find. The discovery of orthogonal bases and local time-frequency analysis opens the door to the world of sparse representation of signals. An orthogonal basis is a dictionary of minimum size that can yield a sparse representation if designed to concentrate the signal energy over a set of few vectors (Mallat, 2009a). The smaller amount of wavelet coefficients reveals the information of signal we are looking for. The generation of those vectors is an approximation of original signal by linear combination of wavelets. For all f in $L^2(R)$,

$$P_j f = P_{j+1} f + < f, \psi_{j,k} > \psi_{j,k}, \tag{31}$$

where $< f, \psi_{j,k} >$ stands for the inner product of f and $\psi_{j,k}$, P_j is the orthogonal projection onto V_j. In orthogonal decomposition, V_j is the subspace which satisfies $\cdots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \cdots$, $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$ and $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$ (Daubechies, 1992).

Thresholding the wavelet coefficients keeps the local regularity of signal. Usually, wavelet thresholding includes three steps (Shim et al., 2001; Zhang et al., 2007):

1. Decomposition. A filter bank of conjugate mirror filters decomposes the discrete signal in a discrete orthogonal basis. The wavelet function $\psi_{j,k}[n]$ and scale function $\phi_{j,k}[n]$ both belong to the orthogonal basis $B = [\{\psi_{j,k}[n]\}_{L<j\le J, 0\le k<2^{-j}}, \{\phi_{J,k}[n]\}_{0\le k<2^{-j}}]$. The scale parameter 2^j varies from $2^L = N^{-1}$ up to $2^J < 1$, where N is the sampling rate of signal X.
2. Thresholding. After decomposition, the threshold is selected. A thresholding estimator in the basis B is written as

$$\tilde{F} = \sum_{j=L+1}^{J} \sum_{k=0}^{2^{-j}} \rho_T(< X, \psi_{j,k} >) \psi_{j,k} + \sum_{k=0}^{2^{-J}} \rho_T(< X, \phi_{J,k} >) \phi_{J,k}, \tag{32}$$

where ρ_T is either a hard threshold or a soft threshold. Normally, the selected threshold is applied on all coefficients except the coefficients contain the lowest frequency energy $< X, \phi_{J,k} >$. This aims to keep the regularity of reconstructed signal. The difference between keeping and not keeping the lowest-frequency approximate coefficients is illustrated by Fig.3. Universal threshold with hard thresholding is used in estimation. The original data in Fig.3(a) has a wide frequency range. It contains both low frequency regular component and high frequency irregular components. Fig.3(c) shows when lowest-frequency part is kept,

the regular component is still included in reconstructed signal. But if the lowest-frequency part is removed, only the high-frequency irregular components left as in Fig.3(d).

3. Reconstruction. After thresholding, all the coefficients are reconstructed to form the denoised signal.

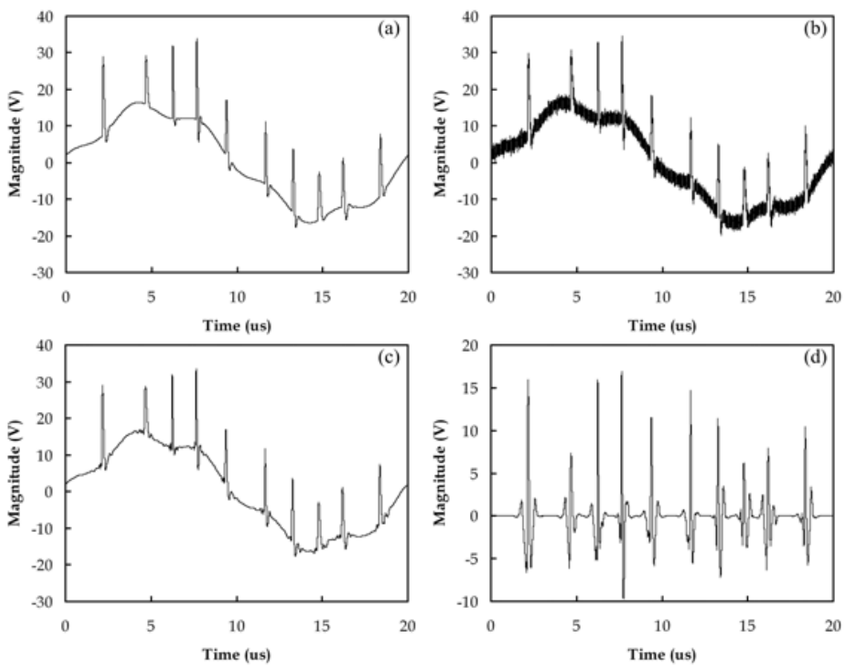


Fig. 3. Difference between keeping lowest-frequency approximates and not keeping it. (a) original data, (b) noisy data (SNR=22.93dB), (c) estimation with $< X, \phi_{J,k} >$ kept (SNR=34.83dB), (d) estimation without $< X, \phi_{J,k} >$ kept (SNR=0.24dB)

Multi-resolution analysis and wavelet packet transform are the most widely employed orthogonal analyses. The wavelet thresholding by using multi-resolution analysis and wavelet packet are introduced in section 4.1 and section 4.2.

4.1 Multi-resolution analysis

Multi-resolution analysis is discrete wavelet transform using series of conjugate mirror filter pairs. The signal f is projected onto a multi-resolution approximation space V_j. This space is then decomposed into a lower resolution space V_{j+1} and a detail space W_{j+1}. The two spaces satisfy $V_{j+1} \perp W_{j+1}$, and $V_{j+1} \oplus W_{j+1} = V_j$ (Daubechies, 1992). The orthogonal basis $\phi_j(t-2^j n)_{n\in\mathbb{Z}}$ of f in V_j is also divided into two new orthogonal bases $\phi_{j+1}(t-2^{j+1}k)_{k\in\mathbb{Z}}$ of V_{j+1}, and $\psi_{j+1}(t-2^{j+1}k)_{k\in\mathbb{Z}}$ of W_{j+1}.

This decomposition process is realized by filtering f by a pair of conjugate mirror filters $h[k]$ and $g[k]=(-1)^{1-k}h[1-k]$. The $h[k]$ and $g[k]$ are also called low pass filter and high

pass filter, respectively. They usually denote the filter banks at reconstruction. At decomposition, the wavelet coefficients are calculated with $\bar{h}[k]$ and $\bar{g}[k]$ where $\bar{h}[k]=h[-k]$ and $\bar{g}[k]=g[-k]$. Accordingly, the coefficients generated by low pass filter and high pass filter are called approximates and detail, respectively (Mallat, 2009b)

$$a_{j+1}[p]=\sum_{k=-\infty}^{+\infty}h[k-2p]a_j[n]=a_j*\bar{h}[2p],\ \text{and}\ d_{j+1}[p]=\sum_{k=-\infty}^{+\infty}g[k-2p]a_j[n]=a_j*\bar{g}[2p]. \tag{33}$$

At the reconstruction,

$$\begin{aligned}a_j[p]&=\sum_{k=-\infty}^{+\infty}h[p-2n]a_{j+1}[n]+\sum_{k=-\infty}^{+\infty}g[p-2n]a_j[n]\\&=\breve{a}_{j+1}*h[p]+\breve{d}_{j+1}*g[p]\end{aligned}. \tag{34}$$

Fig.4 shows the thresholding procedure with multi-resolution analysis.

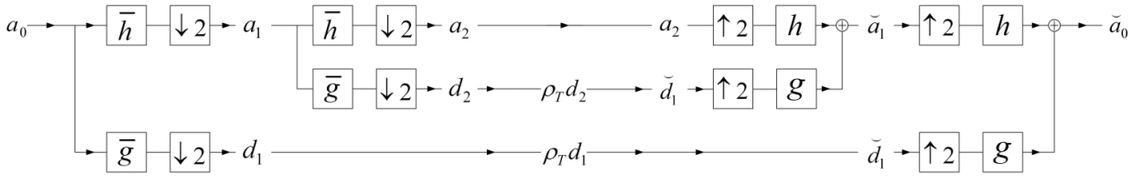


Fig. 4. Thresholding procedure with multi-resolution analysis (the lowest-frequency approximate a_2 is kept)

4.2 Wavelet packet transform

Different time-frequency structures are contained in complex signals. This motivates the exploration of time-frequency representation with adaptive properties. Although similar to multi-resolution analysis, wavelet package can divide the frequency axis in separate interval of various sizes. Its spaces W_j are also divided into two orthogonal spaces. In order to discriminate the detail spaces of wavelet packet from those of multi-resolution analysis, the W_j is represented as W_j^p. Thus, the relation between detail spaces is $W_{j+1}^{2p}\perp W_{j+1}^{2p+1}$, and $W_{j+1}^{2p}\oplus W_{j+1}^{2p+1}=W_j^p$. The orthogonal bases at the children nodes can be represented as $\psi_{j+1}^{2p}(t)=\sum_{k=-\infty}^{+\infty}h[k]\psi_j^p(t-2^jk)$ of W_{j+1}^{2p}, and $\psi_{j+1}^{2p+1}(t)=\sum_{k=-\infty}^{+\infty}g[k]\psi_j^p(t-2^jk)$ of W_{j+1}^{2p+1} (Mallat, 2009c).

Wavelet packet coefficients are computed with a filter bank that is the same as multi-resolution analysis. The wavelet packet transform is an iteration of the two-channel filter bank decomposition presented in section 4.1. At the decomposition, the wavelet coefficients of wavelet packet children d_{j+1}^{2p} and d_{j+1}^{2p+1} are obtained by subsampling the convolutions of d_j^p with low-pass filter $\bar{h}[k]$ and high-pass filter $\bar{g}[k]$:

$$d_{j+1}^{2p}[k] = d_j^p * \bar{h}[2k], \text{ and } d_{j+1}^{2p+1}[k] = d_j^p * \bar{g}[2k]. \tag{35}$$

Iterating the decomposition of coefficients along the branches forms a binary wavelet packet tree with $2^L - 1$ leaves $d_L^n \ (0 \le n \le 2^L - 1)$ at level L. Then, at the reconstruction,

$$d_j^p[k] = \breve{d}_{j+1}^{2p} * h[k] + \breve{d}_{j+1}^{2p+1} * g[k]. \tag{36}$$

The decomposition and reconstruction of wavelet packet transform are illustrated in Fig.5. The thresholding procedure is added before reconstruction.

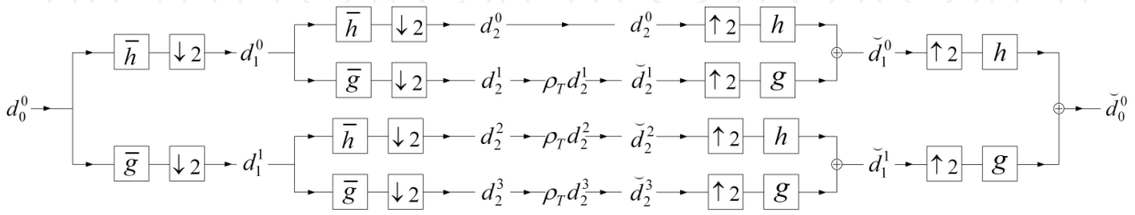


Fig. 5. Thresholding procedure by using wavelet packet transform (the lowest-frequency approximate d_2^0 is kept)

Both multi-resolution analysis and wavelet packet transform are used in estimation of a same noisy signal. The estimations are shown in Fig.6. The coiflet 2 is used to calculate wavelet coefficients and the hard threshold is set as $T = \tilde{\sigma}\sqrt{2\log_e N}$.

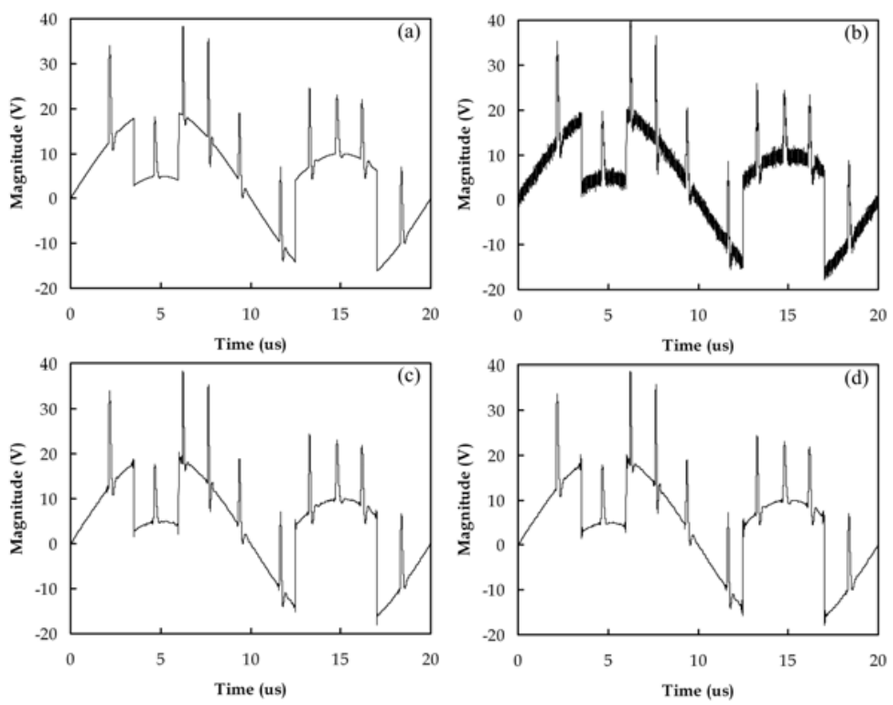


Fig. 6. Estimation with different wavelet transforms, (a) original data, (b) noisy data (SNR=23.56dB), (c) estimation with multi-resolution analysis (SNR=35.3dB), (d) estimation with wavelet packet transform (SNR=33.22dB)

4.3 Noise variance estimation

In threshold estimation discussed in section 3, the variance σ^2 of noise W is an important factor in threshold T. In practical application, the variance is unknown and its estimation is needed. When estimating the variance σ^2 of noise $W[n]$ from the data $X[n]=f[n]+W[n]$, the influence from $f[n]$ must be considered. When f is piecewise regular, a robust estimator of variance can be calculated from the median of the finest-scale wavelet coefficients. Fig.7 depicts the wavelet transforms of three functions: blocks, pulses and heavisine. They are chosen because they often arise in signal processing. It is easy to find the large-magnitude coefficients only occur exclusively near the areas of major spatial activities (Donoho &Johnstone, 1994).

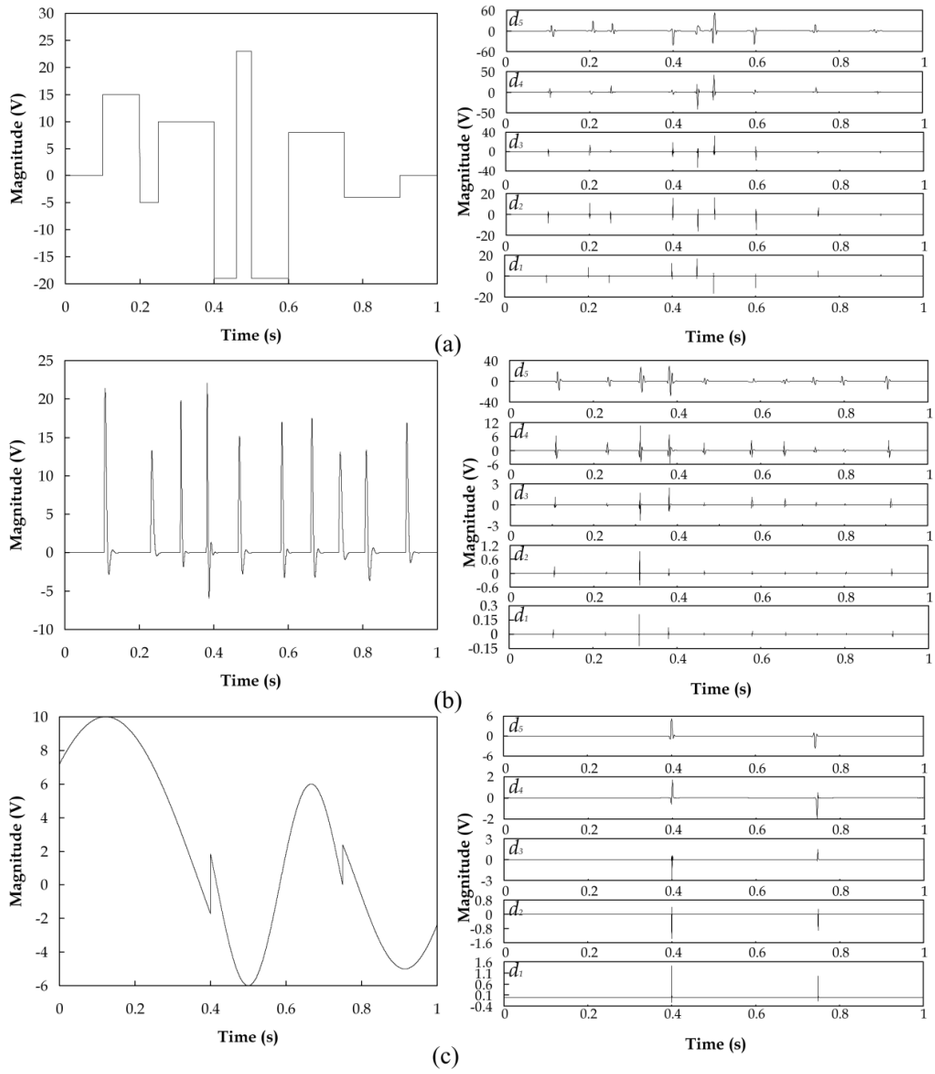


Fig. 7. Three functions and their wavelet transform, (a) blocks (left) and its wavelet coefficients (right), (b) pulses (left) and its wavelet coefficients (right), (c) heavisine (left) and its wavelet coefficients (right)

If f is piecewise smooth, the wavelet coefficients $|< f,\psi_{l,k} >|$ at finest scale l are very small, in which case $< X,\psi_{l,k} > \approx < W,\psi_{l,k} >$. As mentioned in section 3.1, the wavelet coefficients $< W,\psi_{l,k} >$ are still white if W is white. Therefore, most coefficients contribute to noise with variance σ^2 and only a few of them contribute to signal. Then, a robust estimator of σ^2 is calculated from the median of wavelet coefficients $|< W,\psi_{l,k} >|$. Different from mean value, median is independent of the magnitude of those few large-magnitude coefficients related with signal. If M is the median of absolute value of independent Gaussian random variables with zero mean and variance σ_0^2, then one can show that

$$E\{M\} \approx 0.6745\sigma_0 \tag{37}$$

The variance of noise W is estimated from the median M_X of absolute wavelet coefficients $|< W,\psi_{l,k} >|$ by neglecting the influence of f (Mallat, 2009d):

$$\tilde{\sigma} = \frac{M_X}{0.6745} \tag{38}$$

Actually, piecewise smooth signal f is only responsible for a few large-magnitude coefficients, and has little impact on the value of M_X.

4.4 Hard or soft threshold

As mentioned in section 3.2, the estimation can be done with hard and soft thresholding. The estimator $\tilde{F}$ with soft threshold is at least as regular as original signal f since the wavelet coefficients have a smaller magnitude. But this will result in a slight difference in magnitude when comparing estimation with original signal. This is not true if hard threshold is applied. All the coefficients with large-amplitude above threshold T are unchanged. However, because of the error induced by hard-thresholding, oscillations or small ripples are created near irregular points.

Fig.8 shows the wavelet estimation with hard and soft thresholding. The original data consists of a pulse signal and a sinusoidal. It includes both piecewise smooth signal and irregular segments. The wavelet coefficients are calculated with a coiflet 2. The variance σ^2 of white noise is calculated with (38) and the threshold is set to $T = \tilde{\sigma}\sqrt{2\log_e N}$. In Fig.8(c), the hard thresholding removes the noise in the area where the original signal f is regular. But the coefficients near the singularities are still kept. The SNR of estimation with hard thresholding is 36.47dB. Compared with hard thresholding, the magnitude of coefficients with soft thresholding is a little smaller. The soft thresholding estimation attenuates the noise affect at the discontinuities, but it also reduces the magnitude of estimation. The SNR of soft-thresholding estimation reduces to 31.98dB. The lower SNR of soft thresholding doesn't mean poor ability of signal estimation. The two thresholdings are selected in different applications.

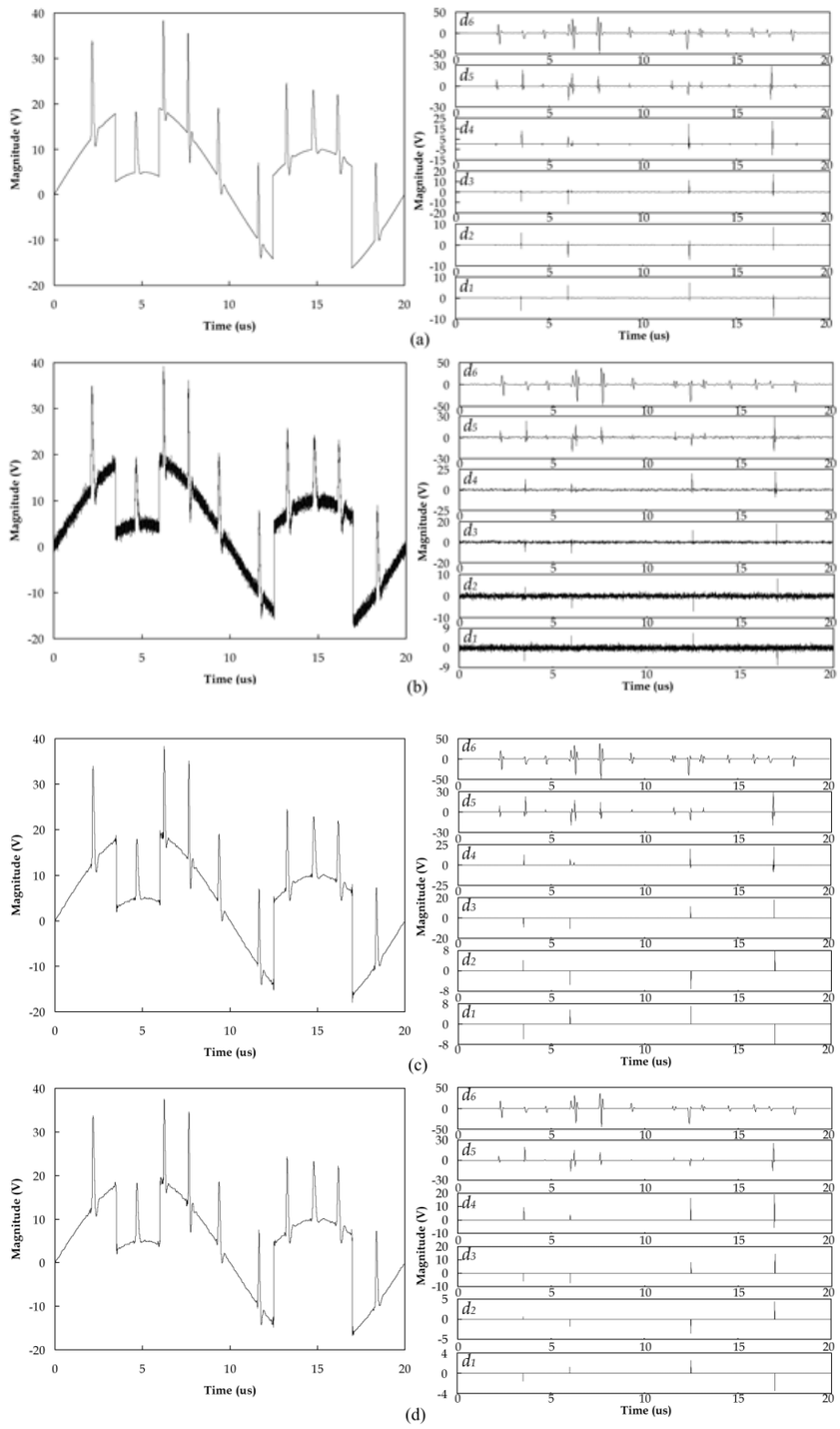


Fig. 8. Wavelet thresholding with hard and soft threshold, (a) original data (left) and its wavelet transform, (b) noisy data (SNR=23.58dB) (left) and its wavelet transform, (c) Estimation with hard thresholding (SNR=36.47dB) (left) and its wavelet coefficients (right), (d) Estimation with soft thresholding (SNR=31.98dB) (left) and its wavelet coefficients (right).

4.5 Fixed or level-dependent variance estimation

If the influence of level is neglected, the estimated variance $\tilde{\sigma}$ of white noise can be set as the estimation of finest scale, or d_1 in Fig.9. As discussed in section 4.3, most wavelet coefficients at finest scale contribute to noise, and only a few of them contributes to signal. The use of fixed estimator $\tilde{\sigma}$ reduces the influence of signal and edge effect in wavelet transform. But when the added noise is no longer white noise, for example, colored Gaussian noises, the noise variance should be estimated level by level (Johnstone & Silverman, 1997). Fig.9 gives the estimation of original signal in Fig.8(a).

In Fig.9, a Gaussian noise is added. Section 4.3 explains how to calculate the threshold value from the wavelet coefficients. Here, the universal threshold $T = \tilde{\sigma}\sqrt{2\log_e N}$ proposed in section 3.3.1 is used. In Fig.9(a), we estimate the noise variance with the median formula in (38) at the finest scale. Only one threshold T is produced. In level-dependent estimation, the estimation of noise variance (38) is done for each scale. That is to say, six scales in Fig.9(b) will generate six estimated variances $\tilde{\sigma}$ and thus six thresholds T. Each threshold is applied on each scale accordingly. The SNR of estimation with level-dependent estimation (36.7dB) is greater than that of fixed estimation (36.4dB).

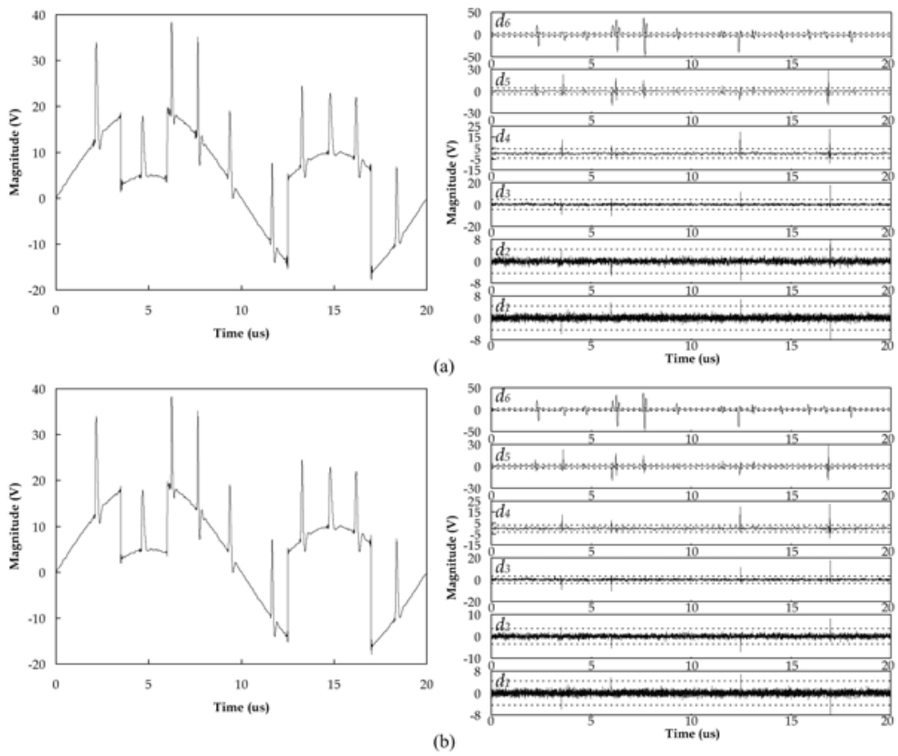


Fig. 9. Wavelet thresholding with fixed and level-dependent variance estimation, (a) Estimation with fixed estimation (SNR=36.4dB) (left) and its wavelet transform (right), (b) Estimation with level-dependent estimation (SNR=36.7dB) (left) and its wavelet transform (right).

5. Selection of optimal wavelet bases

Wavelet thresholding explores the ability of wavelet bases to approximate signal f with only a few non-zero coefficients. Therefore, optimal selection of wavelet bases is an important factor in wavelet thresholding. This depends on the properties of signal and wavelets such as regularity, number of vanishing moments, and size of support.

5.1 Vanishing moments

The number of vanishing moments determines what the wavelet doesn't "see" (Hubbard, 1998). Usually, the wavelet ψ has p vanishing moments if

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0 \quad \text{for } 0 \le k < p \,. \tag{39}$$

This means that ψ is orthogonal to any polynomial of degree $p-1$. Therefore, the wavelet with two vanishing moments cannot see the linear functions; the wavelet with three vanishing moments will be blind to both linear and quadratic functions; and so on. If f is regular and C^k, which means f is p times continuously differentiable function, when $k < p$ then the wavelet can generate small coefficients at fine scales 2^j (Mallat, 2009b). But it is not the higher the better. Too high vanishing moment may miss the useful information in signal, and leave more useless information such as noise. The proper number of vanishing moments is thus important in optimal wavelet selection.

5.2 Size of support

The size of support is the length of interval in which the wavelet values are non-zero. If f has an isolated singularity at t_0 and if t_0 is inside the support of $\psi_{j,k}(t) = 2^{-j/2}\psi_{j,k}(2^{-j}t - k)$, then $< f, \psi_{j,k} >$ may have a large amplitude. If ψ has a compact support of size N, at each scale 2^j there are N wavelets $\psi_{j,k}$ whose support includes t_0 (Mallat, 2009b). In wavelet thresholding application, the signal f is supposed to be represented by a few non-zero coefficients. The support of wavelet should be in a smaller size.

If an orthogonal wavelet ψ has p vanishing moments, its support size must be at least $2p-1$. The Daubechies wavelets are optimal to have minimum size of support for a given number of vanishing moments. When choosing a wavelet, we have to face a trade-off between number of vanishing moments and size of support. This is dependent on the regularity of signal f.

A polynomial function with degree less than 4 is shown in Fig.10. The noisy data and estimations with Daubechies wavelets are also listed. Here, level-dependent threshold is set as $T = \tilde{\sigma}\sqrt{2\log_e N}$. The estimation with wavelet Daubechies 3 whose vanishing moments p is 3 has larger SNR than others.

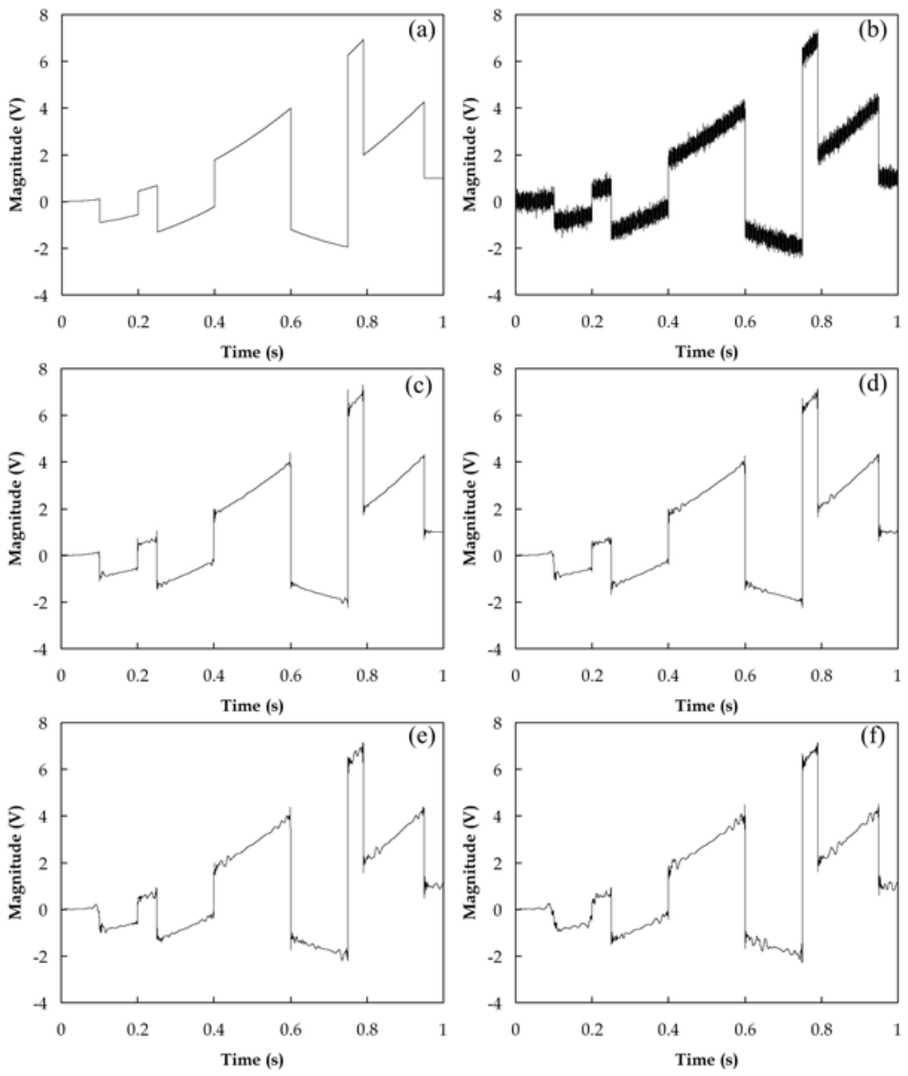


Fig. 10. The estimations by using wavelets with different vanishing moments, (a) original data, (b) noisy data (SNR=22.4dB), (c) estimation with Daubechies 3 (SNR=35.5dB), (d) estimation with Daubechies 7 (SNR=34.6dB), (e) estimation with Daubechies 11 (SNR=33.8dB), (f) estimation with Daubechies 15 (SNR=32.8dB)

5.3 Regularity

The regularity of wavelet induces an obvious influence on wavelet coefficients in thresholding. When reconstructing a signal from its wavelet coefficients $< f, \psi_{j,k} >$, an error ε is added. Then a wavelet component $\varepsilon\psi_{j,k}$ will be added to the reconstructed signal. If ψ is smooth, $\varepsilon\psi_{j,k}$ is a smooth error. For example, in image-denoising, the smooth error is often less visible than irregular errors (Mallat, 2009b). Although the regularity of a function is independent of the number of vanishing moments, the smoothness of some wavelets is related to their vanishing moments such as biorthogonal wavelets.

5.4 Wavelet families

Both orthogonal wavelets and biorthogonal wavelets can be used in orthogonal wavelet transform. Thus, Daubechies wavelets, symlets, coiflets and biorthogonal wavelets are studied in this chapter. Their properties are listed in Table 1 (Mallat, 2009b).

Wavelet name	Order	Number of vanishing moments	Size of support	Orthogonality
Daubechies	$N_{\{1\le N<\infty\}}$	N	$2N-1$	Orthogonal
Symlets	$N_{\{2\le N<\infty\}}$	N	$2N-1$	Orthogonal
Coiflets	$N_{\{1\le N\le 5\}}$	$2N$	$6N-1$	Orthogonal
Biorthogonal wavelets	$N_{d\{1\le N\le 8\}}$ for dec. $N_{r\{1\le N\le 6\}}$ for rec.	N_r	$2N_d+1$ for dec. $2N_r+1$ for rec.	Biorthogonal

Table 1. Information of some wavelet families, 'dec'. is short for decomposition, 'rec'. is short for reconstruction

Choosing the suitable wavelet in wavelet thresholding depends on the features of signal and wavelet properties mentioned in section 5.1, 5.2 and 5.3. For different applications, the optimal wavelets change. For instance, the irregular wavelet Daubechies 2 induces irregular errors in wavelet thresholding of regular signal processing. But it achieves better estimation when applied to estimate transient signal in power system which are often composed by pulses and heavy noises (Ma et al., 2002). As illustrated in Fig.11 and Fig.12, two original datasets are tested with an irregular wavelet Daubechies 2 and a regular wavelet coiflet 3.

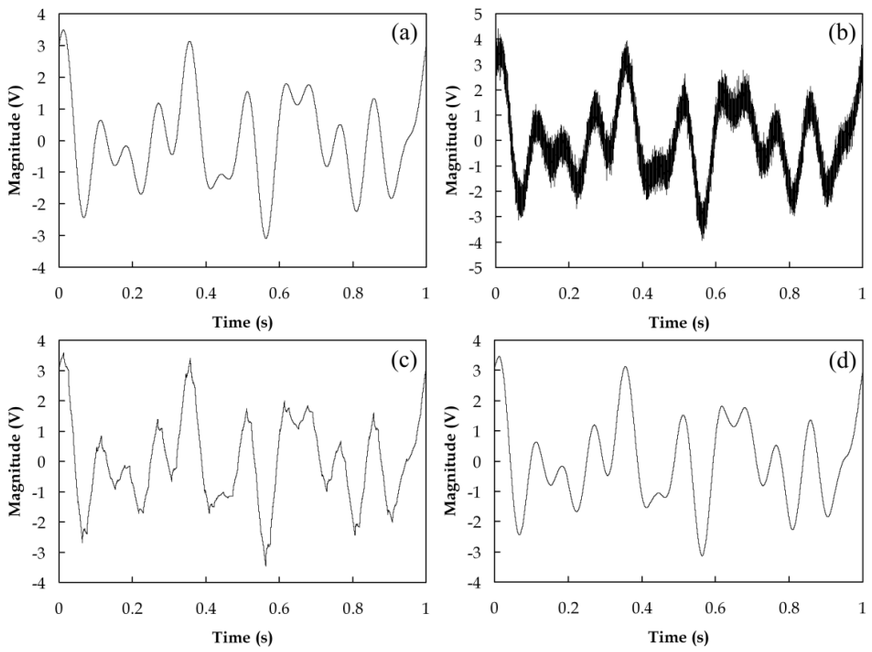


Fig. 11. Estimation of regular data, (a) original data, (b) noisy data (SNR=13.03dB), (c) estimation with 'db2' (SNR=24.75dB), (d) estimation with 'coif3' (SNR=36.96dB)

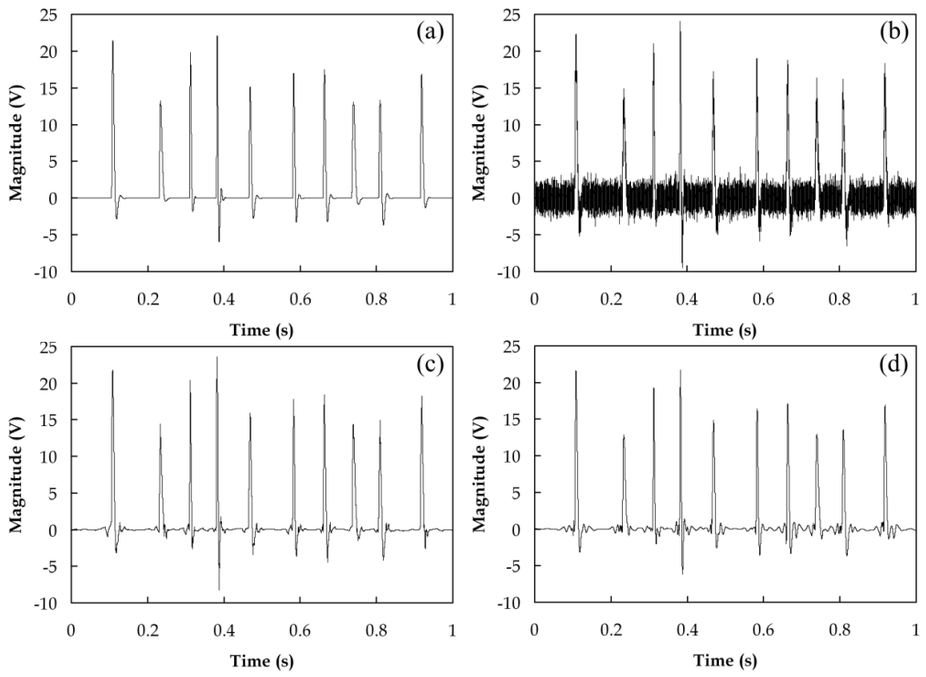


Fig. 12. Estimation of irregular data, (a) original data, (b) noisy data (SNR=10.38dB), (c) estimation with 'db2' (SNR=21.68dB), (d) estimation with 'coif3' (SNR=20.61dB)

6. Conclusion

This chapter focuses on wavelet denoising. It starts with the introduction of two major noise estimation methods: Bayes estimation and Minimax estimation. In orthogonal bases, thresholding is a common method to remove noises. The estimations show that oscillations or ripples will be induced by hard thresholding. Nevertheless, the SNR of estimation with hard thresholding is higher than soft thresholding since the magnitude of coefficients decreases after soft thresholding. Then the thresholds that developed by different noise estimations are proposed. The larger threshold removes more noises but it generates greater estimation risk.

The wavelet denoising methods are usually realized by orthogonal decomposition. The most commonly used orthogonal decompositions are multi-resolution analysis and wavelet packet transform. The influence of wavelet decomposition algorithms, hard or soft thresholdings, and fixed or level-dependent thresholds are studied and compared. For different application, the optimal wavelet thresholding method should be considered carefully.

The wavelet transform is to use a few large magnitude coefficients to represent a signal. The selection of wavelet is another important factor that needs consideration. The properties, for example regularity and degree, of signal should be studied when choosing optimal wavelet that has matching features such as vanishing moments, size of support, and regularity.

7. References

Daubechies, I. (1992). *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, ISBN 0898712742, Philadelphia, Pa.

Donoho D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory*, Vol.41, No. 3, (May, 1995), pp. 613-627, ISSN 00189448

Donoho D. L. & Johnstone I. M. (1994). Ideal Denoising In an Orthonormal Basis Chosen From A Library of Bases. *Comptes Rendus De L Academie Des Sciences Serie I-Mathematique,* Vol. 319, No. 12, (December 1994), pp. 1317-1322, ISSN 0764-4442

Donoho D. L. & Johnstone I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association,* Vol. 90, No. 432, (December 1995), pp. 1200-1224, ISSN 0162-1459

Donoho D. L. & Johnstone I. M. (1998). Minimax estimation via wavelet shrinkage. *Annals of statistics*, Vol. 26, No. 3, (June 1998), pp. 879-921, ISSN 00905364

Hubbard, B. B. (1998). *The world according to wavelets: the story of a mathematical technique in the making,* A. K. Peters, ISBN 1568810725, Wellesley, Mass

Johnstone I. M. & Silverman B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the royal statistical society: series B (statistical methodology)*, Vol. 59, No.2, (May, 1997), pp. 319-351, ISSN 13697412

Lehmann, E. L. & Casella G. (1998). *Theory of point estimation*, Springer, ISBN 0-387-98502-6, New York

Ma, X., Zhou, C. & Kemp, I. J. (2002). Automated wavelet selection and thresholding for PD detection. *IEEE Electrical Insulation Magazine*, Vol. 18, No. 2, (August, 2002), pp. 37-35, ISSN 0883-7554

Mallat, S. G. (2009). Sparse Representations, In: *A wavelet tour of signal processing : the Sparse way,* Mallat, S. G., pp. 1-30, Elsevier/Academic Press, ISBN 13:978-0-12-374370-1, Amsterdam,Boston

Mallat, S. G. (2009). Wavelet Bases, In: *A wavelet tour of signal processing : the Sparse way,* Mallat, S. G., pp. 263-370, Elsevier/Academic Press, ISBN 13:978-0-12-374370-1, Amsterdam,Boston

Mallat, S. G. (2009). Wavelet Packet and Local Cosine Bases, In: *A wavelet tour of signal processing : the Sparse way,* Mallat, S. G., pp. 377-432, Elsevier/Academic Press, ISBN 13:978-0-12-374370-1, Amsterdam,Boston

Mallat, S. G. (2009). Denoising, In: *A wavelet tour of signal processing : the Sparse way,* Mallat, S. G., pp. 535-606, Elsevier/Academic Press, ISBN 13:978-0-12-374370-1, Amsterdam,Boston

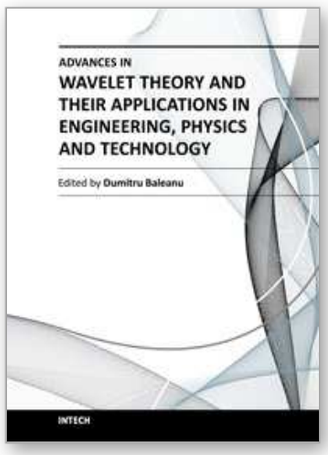
Shim I., Soraghan J. J. & Siew W. H. (2001). Detection of PD utilizing digital signal processing methods. Part 3: Open-loop noise reduction. *IEEE Electrical Insulation Magazine*, Vol. 17, No. 1, (February, 2001), pp.6-13, ISSN 0883-7554

Stein C. M. (1981). Estimation of the Mean of a Multivariate Normal-Distribution. *Annals of Statistics,* Vol. 9, No. 6, (November 1981), pp. 1317-1322, ISSN 0090-5364

Zhang H., Blackburn T. R., Phung B. T. & Sen D. (2007). A novel wavelet transform technique for on-line partial discharge measurements. 1. WT de-noising algorithm. *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 14, No. 1, (February, 2007), pp. 3-14, ISSN 1070-9878

IntechOpen

IntechOpen



Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology

Edited by Dr. Dumitru Baleanu

ISBN 978-953-51-0494-0
Hard cover, 634 pages
Publisher InTech
Published online 04, April, 2012
Published in print edition April, 2012

The use of the wavelet transform to analyze the behaviour of the complex systems from various fields started to be widely recognized and applied successfully during the last few decades. In this book some advances in wavelet theory and their applications in engineering, physics and technology are presented. The applications were carefully selected and grouped in five main sections - Signal Processing, Electrical Systems, Fault Diagnosis and Monitoring, Image Processing and Applications in Engineering. One of the key features of this book is that the wavelet concepts have been described from a point of view that is familiar to researchers from various branches of science and engineering. The content of the book is accessible to a large number of readers.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Guomin Luo and Daming Zhang (2012). Wavelet Denoising, Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology, Dr. Dumitru Baleanu (Ed.), ISBN: 978-953-51-0494-0, InTech, Available from: http://www.intechopen.com/books/advances-in-wavelet-theory-and-their-applications-in-engineering-physics-and-technology/wavelet-denoising

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen