# ELIMINATION OF THE MEASUREMENT NOISE BY USING WAVELET TRANSFORMATIONS

*Marek Kupczak*

Department of control systems and instrumentation
Faculty of Mechanical Engineering
VŠB-TU Ostrava

**Abstract**

This contribution describes a method for ideal de-noising. This method is based on a choice of an optimal basis from a library of wavelet orthogonal bases. This is the process of minimizing of the cost functional of entropy in contribution with shrinkage method. In particular, the chosen shrinkage methods are the estimation of standard deviation $\sigma$ and of Donoho-Johnstone universal threshold $\lambda$ for the wavelet coefficients. The chosen entropy is Shannon entropy and is defined by the discrete distribution of squared wavelet coefficients. The Donoho thresholding rule belongs to the simplest non-linear shrinkage technique.
The result of this ideal de-noising is presented on one example. The algorithm, selecting an optimal basis from a library of wavelet bases for ideal de-noising, was created with using MATLAB and The Wavelet Toolbox.

## 1. Representation of the observations

Suppose we have observations $y_i,\ i = 1, \ldots, N$, usually $N = 2^m$. It is natural that these observations are corrupted by measurement noise. So, observations can be represented by (1) or by the vector form (2)

$$y_i = s_i + \sigma \cdot e_i, \quad i = 1, \ldots, N \tag{1}$$

$$\mathbf{y} = \mathbf{s} + \sigma \cdot \mathbf{e} \tag{2}$$

where $\mathbf{s} = \left( s_1, \ldots, s_N \right)$ is the signal of interest and $\mathbf{e} = \left( e_1, \ldots, e_N \right)$ is white noise with normal (Gaussian) probability distribution $\mathrm{N}\left(0, \sigma^2\right)$, $e_i$'s are zero-mean random variables. We wish to recover $\mathbf{s}$ with small risk or mean-squared error (MSE) defined by (3)

$$\boldsymbol{R}\left( \hat{\boldsymbol{s}}, \boldsymbol{s} \right) = \frac{1}{N} \boldsymbol{E} \, \| \hat{\boldsymbol{s}} - \boldsymbol{s} \|_{\ell_2}^2 \tag{3}$$

where $\hat{s}(\boldsymbol{y}) = \left(s_1, \ldots, s_N\right)$ is a discrete estimator of $\mathbf{s}$.

In the recent time, it is very popular to use a fixed wavelet orthogonal basis for noise removal. This process known as thresholding scheme can be divided into three steps:

1. Transform of observations into the wavelet basis
2. Apply thresholding
3. Return to the original basis

Success of such a de-noising scheme depends on the chosen basis in which we will recover the signal $\mathbf{s}$. Since a given signal may be recovered well in one basis and not in others, others signals may not be recovered as well in the same basis.

## 2. Discrete wavelet transformations

Discrete wavelet transformations map discrete data $y_i,\ i=1,\dots,N$ from the time domain to the time-frequency (wavelet) domain $Y_i, i=1,\dots N$. The result is a vector of the same size. Wavelet transformations are linear and they can be defined by matrices of dimension $N\times N$. When the matrix is orthogonal, the corresponding transformation is a rotation in $\mathbf{R}^N$ in which the data represent coordinates of a point.

Any function $f\in\ell_2(\mathbf{R})$ can be represented as (4)

$$f(t)=\sum_{j,k} d_{jk}\,\psi_{jk}(t) \tag{4}$$

($\psi$ is the wavelet function, $\varphi$ is the scaling function)

This unique representation corresponds to a multiresolution decomposition $\boldsymbol{l_2}(\mathbf{R})=\bigoplus_{j=-\infty}^{\infty} W_j$. Also, for any fixed $j_0$ the decomposition $\boldsymbol{l_2}(\mathbf{R})=V_{j0}\oplus\oplus_{j=j_0}^{\infty} W_j$ corresponds to the representation (5)

$$f(t)=\sum_k c_{j_0 k}\,\varphi_{j_0 k}(t)+\sum_{j\ge j_0}\sum_k d_{jk}\,\psi_{jk}(t) \tag{5}$$

The first sum in (5) is an orthogonal projection $\boldsymbol{P}_{j_0}$ of $\boldsymbol{f}$ on scaling functional spaces $\boldsymbol{V}_{j_0}$ while the other sums are orthogonal complements of wavelet functional spaces $\boldsymbol{W}_j$.

The decomposition algorithm (the cascade algorithm and dyadic wavelets) can be described in the matrix notation as follows:

Let the length of the input signal $\boldsymbol{f}$ be $N=2^m$, let $\boldsymbol{h}=\{h_s, s\in\boldsymbol{Z}\}$ be the wavelet filter of length $K$, i.e. only $K$, entries of $\boldsymbol{h}$ are different from zero.

Denote by $\mathbf{H}_k$ a circulant matrix of size $\left(2^{m-k}\times 2^{m-k+1}\right)$, $k=1,2,\dots$ with entries (6)

$$h_s,\ s=(K-1)+(i-1)-2\,(j-1)\ \mathrm{mod}\,2^{m-k+1} \tag{6}$$

as the position $(i,j)$. So, $i$th row is 1st row circularly shifted to the right by $2(i-1)$ units. This circularity is a consequence of using the modulo operator in (6). Similarly, we define a matrix $\mathbf{G}_k$ by using the filter $\boldsymbol{g}$. Coefficients of $\boldsymbol{g}$ and $\boldsymbol{h}$ are related by $g_s=(-1)^s h_{K+1-s}$. The total matrix is $\mathbf{W}_k=\begin{pmatrix}\mathbf{H}_k\\ \mathbf{G}_k\end{pmatrix}$.

Now, we can describe observations in the wavelet domain by (7) or by the vector form (8)

$$Y_i=\theta_i+\sigma\cdot z_i,\quad i=1,\dots,N \tag{7}$$

$$\mathbf{Y}=\boldsymbol{\theta}+\sigma\cdot\mathbf{z} \tag{8}$$

with $\mathbf{Y}=(Y_1,\dots,Y_N)=\mathbf{W}\cdot\mathbf{y}$, $\boldsymbol{\theta}=(\theta_1,\dots,\theta_N)=\mathbf{W}\cdot\mathbf{s}$ and $\mathbf{z}=(z_1,\dots,z_N)=\mathbf{W}\cdot\mathbf{e}$. Due to orthogonality of $\mathbf{W}$, $\mathbf{z}\sim\mathrm{N}(0,\sigma^2)$ and $\mathrm{MSE}(\hat{\boldsymbol{s}},\boldsymbol{s})=\mathrm{MSE}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})$, moreover it's very simple to find inverse matrix because $\mathbf{W}^{-1}=\mathbf{W}^T$.

The first step of decomposition is defined by $\mathbf{Y}=\begin{pmatrix}\mathbf{cA_1}\\ \mathbf{cD_1}\end{pmatrix}=\mathbf{W}\cdot\mathbf{y}$ where $\mathbf{cA_1}$ is the approximation coefficient vector and $\mathbf{cD_1}$ is the detail coefficient vector. The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken into many lower-resolution components. This process can be shown in Fig. 1.

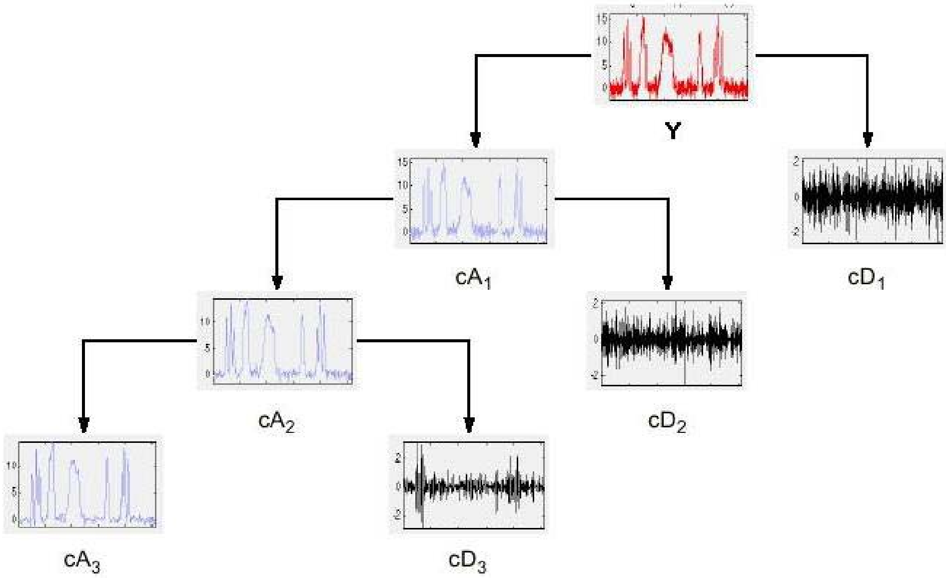


Figure 1 – Multiresolution decomposition

### 3. Thresholding

The thresholding is the simplest wavelet non-linear shrinkage technique. It is common for all thresholding rules to set to 0 the coordinates of a vector $\mathbf{Y}$, which is subjected to thresholding, if they are smaller in absolute value than a fixed non-negative number - the threshold $\lambda$. Depending on how the coordinates of $\mathbf{Y}$ are processed when they are larger than $\lambda$ one can define different thresholding policies. The two most common thresholding policies are hard and soft. The analytic expressions for the hard- and soft-thresholding rules are (9) and (10)

$$\delta^h(Y_i,\lambda)=Y_i\mathbf{1}\left(|Y_i|>\lambda\right) \quad (9)$$

$$\delta^s(Y_i,\lambda)=\left(Y_i-\mathrm{sgn}(Y_i)\cdot\lambda\right)\mathbf{1}\left(|Y_i|>\lambda\right) \quad (10)$$

for $\lambda \geq 0$, $\mathbf{Y}=(Y_1,\ldots,Y_N)\in\mathbf{R}$. The rules are depicted in Fig. 2.

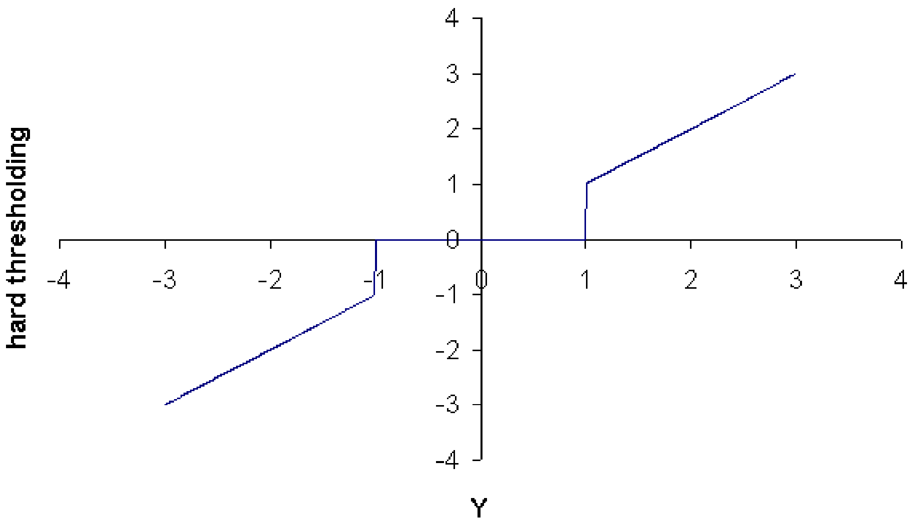


Figure 2 – Hard thresholding rule for $\lambda = 1$

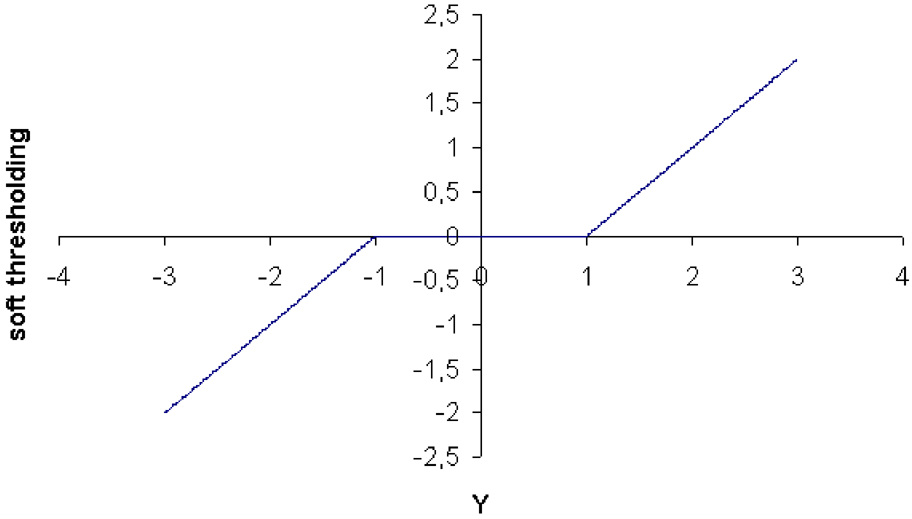


Figure 3 – Soft thresholding rule for $\lambda = 1$

The optimal thresholding rules are obtained by choosing the threshold $\lambda = \sigma$. Of course, $\sigma$ is generally not known, and the optimal risk remains unattainable. Assume the model (7), respectively (8). Suppose we have an oracle that tells us which of $\theta_i^{'}s$ are close to 0. The oracle suggests only two actions: "keep" or "kill" the observation with the index $i$. So, the $\hat{\theta}_i$ suggested by the oracle is $\hat{\theta}_i = Y_i \partial_i$, where $\partial_i = 0$ or 1. The ideal $\partial$ in this case is given coordinate-wise by $\partial = \mathbf{1}\left(\left|Y_i\right| > \sigma\right)$. The diagonal projection (DP) estimator is $\hat{\boldsymbol{\theta}} = \left\{Y_i \, \boldsymbol{1}\left(\left|Y_i\right| > \sigma\right), i = 1, \ldots, N\right\}$, and its risk is expressed by

$$R(DP, \boldsymbol{\theta}) = E\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_{\ell_2}^2 = \sum_i \min\left(\left|\theta_i\right|^2, \sigma^2\right) \tag{11}$$

which can be readily derived by discussing the two possible cases: $\left|Y_i\right| < \sigma$ and $\left|Y_i\right| \geq \sigma$. The oracular risk (11) is ideal and unattainable if $\sigma$ is not known. However, it is useful as a benchmark for evaluating other rules.
Donoho and Johnstone have shown that the risk of the soft-thresholding rule with an universal threshold $\lambda = \sigma\sqrt{2\log N}$ is close [up to a multiple $\left(1 + 2\log N\right)$] to the oracular risk $R\left(DP, \boldsymbol{\theta}\right)$.

$$R\left(\hat{\boldsymbol{\theta}}^*, \boldsymbol{\theta}\right) \leq \left(2\log N + 1\right) \cdot \left[\sigma^2 + R\left(DP, \boldsymbol{\theta}\right)\right] \tag{12}$$

## 4. Choice of an ideal wavelet basis from a library

Suppose we have a library **L** consisting of finitely many orthogonal bases $\mathbf{L} = \left\{\mathbf{B}_1, \ldots, \mathbf{B}_L\right\}$, where $\mathbf{B} = \left\{g_i, \, i = 1, \ldots, N\right\}$ denotes a set of real orthogonal *N*-dimensional vectors which form columns of a matrix of discrete wavelet transform of *N*-dimensional data vectors. The created library **L** can contain such wavelet bases as bases from Daubechies' family or biorthogonal family. The best ideal risk in any basis in the library is (13)

$$R^*\left(\mathbf{s}, \mathbf{L}\right) = \min_{B \in \boldsymbol{L}} R\left(\mathbf{s}, \mathbf{B}\right) \tag{13}$$

of course this risk is achievable only with the aid of a basis oracle, which selects for us the basis achieving the optimum; a coordinate oracle informing which coordinates in that basis are worth estimating is also necessary.
Selecting an ideal basis by using the best ideal risk (13), respectively by using the oracular risk (11) and (12), is practically unavailable, because these risks are functionally depend on the vector $\mathbf{s}$, respectively on the vector $\boldsymbol{\theta}$. Of course, these vectors are unknown. We want to find them. So, it means that we need another criteria upon which the bases will be judged.
Such criteria are usually expressed in the form of a cost functional **E**. One of these cost functionals is Shannon entropy defined by (14)

$$\mathbf{E}(\mathbf{Y},\mathbf{B}) = -\sum_{i=1}^{N} P_i^{(\mathbf{y})} \cdot \ln\left(P_i^{(\mathbf{y})}\right) \tag{14}$$

$P$ in this notation means a probability and it's defined as the discrete distribution of squared wavelet coefficients $\mathbf{Y}_j$ by (15)

$$P_i^{(\mathbf{y})} = \frac{\left| Y_{ji}^{(\mathbf{y})} \right|^2}{\left\| Y_j^{(\mathbf{y})} \right\|^2} \tag{15}$$

where $\left\| Y_j^{(\mathbf{y})} \right\|^2 = \sum_{i=1}^{N} \left| Y_{ji}^{(\mathbf{y})} \right|^2$ is the norm of the vector containing wavelet coefficients. For us there are only approximation coefficient vectors extracted by decomposition (*j* is the number of the decomposition level).

Now we can find the best, respectively the optimal orthogonal basis according to Shannon entropy (14), and moreover we find out the optimal decomposition level with the minimal entropy.

$$\hat{\mathbf{B}} = \arg\min_{\mathbf{B} \in \mathbf{L}} \mathbf{E}(\boldsymbol{Y},\mathbf{B}) \tag{16}$$

## 5. Calculation of the standard deviation

When we have estimated the best basis from a library and the optimal decomposition level with the minimal entropy, we can estimate the value of standard deviation $\hat{\sigma}$. We can propose that the useful signal $\mathbf{s}$ is mainly low-frequency and the noise is mainly concentrated on the first detail level. We can define the following inequality (17) with initialisation $M = 0$

$$\frac{\left| Y_{M+1}^{(\text{sort}\,\mathbf{y})} \right|^2}{\sum_{k=M+1}^{N} \left| Y_k^{(\text{sort}\,\mathbf{y})} \right|^2} \le \frac{2 \cdot \ln(N-M)}{N-M} \tag{17}$$

where $Y_k^{(\text{sort}\,\mathbf{y})}$, $k = 1,\ldots,N$ is the sequence of wavelet coefficients sorted in the order of decreasing of their absolute values: $\left| Y_k^{(\text{sorted}\,\mathbf{y})} \right| \ge \left| Y_{k+1}^{(\text{sorted}\,\mathbf{y})} \right|$. (The first member is the wavelet coefficient with the maximum absolute value.) The inequality (17) helps us to find the beginning of the detail part, respectively the end of the approximation part. Now, it's possible to separate the detail part of the decomposition on the first level.

The decision rule (17) is based on the fact resulting from asymptotic distribution of independent Gaussian random values:

$$\textit{If} \quad X(t) \sim N\left(0,\sigma^2\right) \;\Rightarrow\; \lim_{N\to\infty} P\left\{ \frac{\max_{1\le t\le N} |X(t)|^2}{\|X\|^2} \le \frac{2\ln N}{N} \right\} = 1 \tag{18}$$

where $P$ denotes the probability.

Finally, we can already compute the robust estimate of standard deviation of the noise

$$\hat{\sigma} = median\left\{ \left| Y_n^{(\text{sorted}\,y)} \right| \right\}, \quad n = M, M+1, \ldots, N \tag{19}$$

In case when the value of $M$ is greater than the maximal length of the input signal, we cannot use the equation (19) to calculating of the standard deviation because we cannot find the beginning of the first detail level. So, we can use equation (20)

$$\hat{\sigma} = median\left\{ \left| \boldsymbol{cD}_{\boldsymbol{1}}^{(\text{sorted})} \right| \right\} \tag{20}$$

## 6. Designed algorithm

Designed algorithm realizing the optimal basis choice from a library was created by using MATLAB. Name of he created function is *ideal_denoising*. This function has three input arguments:

| | |
|---|---|
| 1. y | input vector containing sample data of observations $\boldsymbol{y}$ |
| 2. max_level | maximal level of decomposition |
| 3. library | string containing names of wavelet bases |

and five output arguments:

| | |
|---|---|
| 1. Emin | Minimal Shannon entropy for the best basis from a library |
| 2. type | type of the best wavelet basis from a library |
| 3. level | minimal decomposition level for the best wavelet basis |
| 4. sigma | the calculated robust estimate of standard deviation of the noise |
| 5. smooth | vector containing the smoothed function according to the best wavelet basis from ibrary, the optimal decomposition level and the hard thresholding rule with the niversal (Donoho-Johnstone) threshold $\lambda$ |

```
function [Emin,type,level,sigma,smooth] = ideal_denoising(y,max_level,library)

ly = length(y);
a = y;
minimum = 0;

% Searching for the minimal Shannon entropy
% of the all bases of the library
for basis = 1:1:length(library)
   for i=1:1:max_level
      [a,d] = dwt(a,library{basis});
      E(basis,i) = wentropy(a,'shannon');
      if (E(basis,i) >= minimum)
         level(basis,1) = i-1;
         break;
      else
         minimum = E(basis,i);
      end
   end
   a = y;
   minimum = 0;
end
E = E';
% Searching for the best wavelet basis
% with the minimal Shannon entropy
% and the appropriate decomposition level
absE = abs(E);
[i,j] = find(absE == max(max(absE)));
Emin = E(i,j);
level = i;
type = library{j};

% Searching for the first detail part
[a,d] = dwt(y,library{j});
coef = [a;d];
abscoef = sort(abs(coef));
k=0;
for i=1:1:length(abscoef)
   sortedcoef(i) = abscoef(end-k);
   k = k+1;
end
sortedcoef = sortedcoef';
N = length(sortedcoef);
for M = 0:1:N
   lnum = abs(sortedcoef(M+1)).^2;
   lden = sum(abs(sortedcoef((M+1):N)).^2);
   left_side = lnum/lden;
   right_side = 2*log(N-M)/(N-M);
   [left_side right_side];
   if (left_side > right_side)
      break;
   end
```

```
end
if (M==0)
   M=1;
end

% Computing of the standard deviation
if ( M < ly)
   sigma = median(abs(sortedcoef(M:N)))/0.6745;
else
   sigma = median(abs(sort(d)))/0.6745;
end

% Hard thresholding on the optimal decomposition
% level with the threshold lambda and by using the best wavelet basis of a library
[C,L] = wavedec(y,level,type);
lambda = sigma*sqrt(2*log(ly));
smooth = wdencmp('gbl',y,type,level,lambda,'h',1);
```

### 7. Creating of the new library

It is very simple to create a new library or add a new wavelet basis to the existing library. We can use a command *save*. On the following example is shown how to create a new library containing Daubechies' wavelet bases from the Wavelet Toolbox and how to add a new basis:

```
Daubechies = {'db1','db2'};
Daubechies{end+1} = 'db3';
save Daubechies
```

It is also very simple to load an existing library by using the command *load*:
load Daubechies

### 8. Case study

Now it's shown a result of using the designed algorithm. The signal (composed of three sine functions with the amplitudes and frequencies in this order: $A_1 = 100$, $f_1 = 1\ kHz$; $A_2 = 50$, $f_2 = 0{,}5\ kHz$; $A_3 = 20$, $f_3 = 0{,}2\ kHz$ ) is corrupted by the white noise of the variance approximately $\sigma^2 = 98{,}7778$, respectively by standard deviation $\sigma = 9{,}9387$ . The sampling frequency is $f_S = 0{,}0001\ Hz$ . Created library *Daubechies* contains 20 Daubechies' wavelet bases from 'db1' to 'db20' (for more see The Wavelet Toolbox Manual). Maximal decomposition level is 20.

The algorithm selected as the best wavelet basis 'db20', decomposition level 20 with the minimal Shannon' entropy $E = -7{,}11 \cdot 10^{12}$ . The estimated standard deviation is $\hat{\sigma} = 9{,}9935$ and it is very close to the real standard deviation of the noise. The canculated value of the threshold $\lambda$ is 42, 8916.

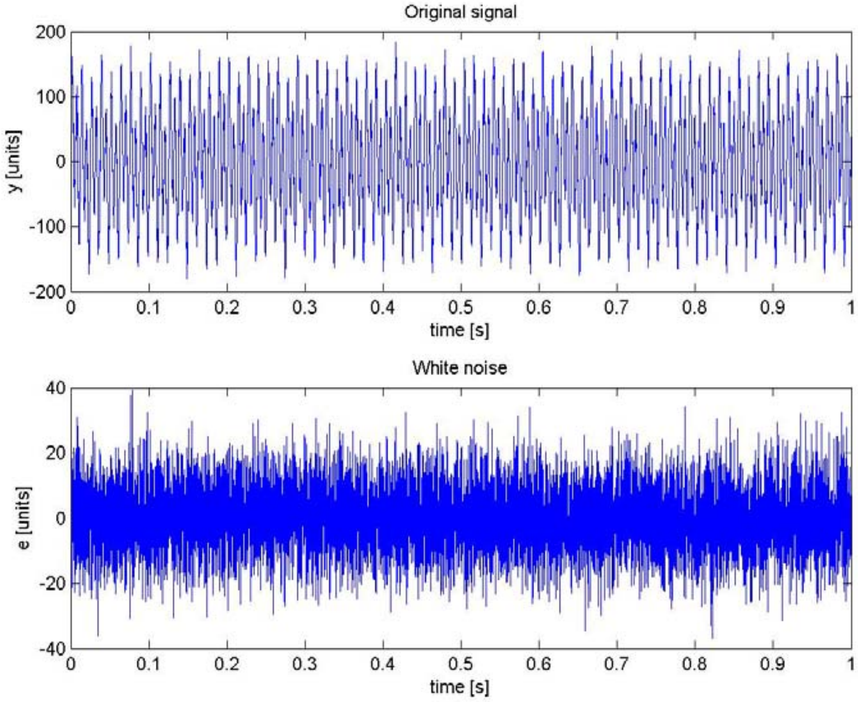


Figure 4 – Original signal and white noise

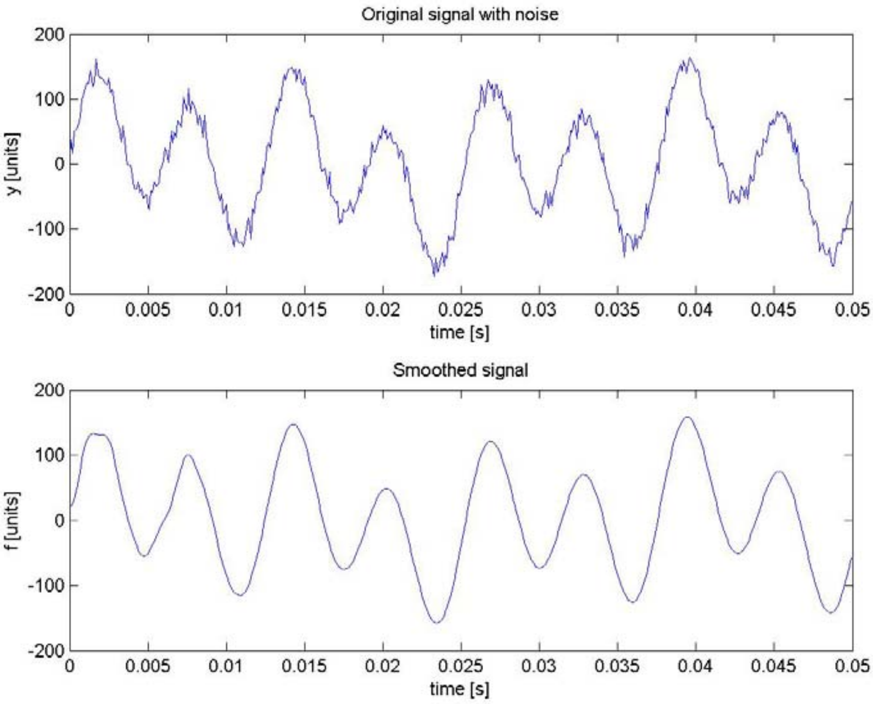


Figure 5 – Original signal with noise and smoothed signal (detail)

**9. Conclusion**

This contribution dealt with the problem that we have a lot of the wavelet bases to choose for signal processing. It was resolved by choosing the ideal basis from the created library based on Shannon' entropy. Further, the measurement noise was cancelled by tresholding. The value of threshold $\lambda$ was canculated by estimating the value of the standard deviation $\sigma$ of this noise.

## References

ČASTOVÁ, N., KALÁB, Z., LYUBUSHIN, A., *Discrete wavelet transform analysis of digital records of seismic signals*. In 1st International Conference APLIMAT 2002. Bratislava: Slovak University of Technology, 2002, 6p. ISBN 80-227-1654-5

DONOHO, D., JOHNSTONE, I. Ideal de-noising in an orthonormal basis chosen from a library of bases. Stanford: Department of Statistics, 1994. 8p.

DONOHO, D., JOHNSTONE, I. Ideal spatial adaptation by wavelet shrinkage. Stanford: Biometrika, 1994. 425-455p.

MISITI, M., MISITI, Y., OPPENHEIM, G., POGGI, J. *Wavelet Toolbox. For Use with MATLAB*. Natick: The MathWorks, 1996. 400p.

VIDAKOVIC, B. *Statistical modeling by wavelets*. New York: John Wiley & Sons, 1999. 400p. ISBN 0-471-29365-2.

## Contact

Ing. Marek Kupczak, Katedra ATŘ-352, VŠB-TU Ostrava, 17. listopadu, Ostrava – Poruba, 708 33
marek.kupczak.fs@vsb.cz