Konrad BOJAR
Industrial Research Institute for Automation and Measurements

Trend extraction from noisy discrete signals by means of singular spectrum analysis and morphological despiking

Abstract. *In real-world applications we deal with electric signals subject to additive stochastic noise of unknown, non-stationary distribution and strong additive shot noise. At the time of measurement the waveform of the signal is often not known what excludes specific denoising schemes. To remedy this, we propose a self-adapting algorithm for effective spike removal and trend extraction without losing essential properties of the signal waveform. Applicability of this algorithm was tested using stress measurements taken in strong electromagnetic fields of wide spectrum.*

Streszczenie. *W rzeczywistych zastosowaniach mamy do czynienia z sygnałami zakłóconymi addytywnym szumem stochastycznym o nieznanym rozkładzie i silnym szumem śrutowym. Proponujemy samoadaptujący algorytm usuwania impulsów i ekstrakcji trendu bez utraty charakteru przebiegu. Algorytm przetestowano na pomiarach tensometrycznych, zarejestrowanych w silnym polu elektromagnetycznym o szerokim widmie.* ***(Ekstrakcja trendu z zaszumionych sygnałów dyskretnych za pomocą analizy widma osobliwego i morfologicznego usuwania impulsów)***

Keywords: morphological filtering, singular spectrum analysis, shot noise, non-stationary noise.
Słowa kluczowe: filtracja morfologiczna, analiza widma osobliwego, szum śrutowy, szum niestacjonarny.

Introduction

It is often that the measured signal contains strong parasitic components that severely impede any further analysis and conceal its required content. Such a situation poses no problem as far as the noise and the useful component of the signal have deterministic or stochastic parameters that can be somehow determined on-the-fly or computed beforehand. Once these parameters are known, a specific denoising procedure can be chosen from a wide range of well-established methods. The problem turns way more complicated when variability of the signal is hard to describe in mathematical terms, and the noise is non-stationary and produces outliers which often group together creating various formations. Hardness of this task resembles this of the blind signal separation problem [1]. In this paper we prove that the solution is possible if only the following is assumed:

a) The noise is of zero mean and can be split into a shot noise of signal-to-noise ratio (SNR) substantially greater that unity and a stochastic noise of substantially lower amplitude that the shot noise. It is important that we do not impose any condition on the SNR of the stochastic component of the noise and there are no other assumptions on its probability density function (PDF).

b) The extreme values of the shot noise are seen as sharp peaks and dips and the waveforms of single shots are of pairwise similar morphology.

Using only these two assumptions we can build a relatively simple and self-adapting algorithm that effectively removes the noise described above. This algorithm combines two already known methods: morphological filtering and singular spectrum analysis. One can argue that the same result could be achieved by spectral methods [2, 3] or filtering based on auto-regressive moving average (ARMA) models [4] or its variants (like seasonal ARMA models [5]), but to the very best of the author's knowledge, when a) and b) only is assumed, these methods are likely to fail.

The paper is organized as follows. At first we discuss principles of 1D morphological operations which will be used to remove the shot noise. Next, we outline basic principles of the singular spectrum analysis which will be applied to recover the trend irrespectively of the noise PDF. Then, we present the trend extraction algorithm itself in terms of a sequence of mathematical operations. To allow prompt application by anyone interested, we give a code of the algorithm in MATLAB. At last, we discuss a sample application of this algorithm and present obtained results.

Principles of mathematical morphology for 1D signals

Let f be a signal and let b be an another signal called structuring element. We define the dilation and the erosion of f by b to be [6]

(1) $$d(f,b)(x) = \boldsymbol{max}_y(f(y)+b(x-y)),$$

(2) $$e(f,b)(x) = \boldsymbol{min}_y(f(y)-b(y-x)),$$

respectively. In a puristic formulation, the dilation and erosion operations are defined using the supremum and infimum operators instead of maximum and minimum, respectively. However, in our case of a finite, discrete set of sampling points these operators coincide. It must be noted that the above morphological operations, as well as the operations that will be defined shortly, are not restricted to the 1D case; they work in the same manner for any number of dimensions. It is much easier to understand these operations in a 2D case of binary images. In such a case f and b can be interchangeably interpreted as sets or characteristic functions of sets. Then, the dilation operation is simply the locus of points covered by all instances of the structuring element b with the center inside f, and the erosion operation can be understood as the set of centers of all instances of the structuring element b lying completely inside f. The dilation and erosion as such are not very useful in our scheme of signal denoising. Therefore, we define the morphological closing and opening operations by

(3) $$f \bullet b = e(d(f,b),b),$$

(4) $$f \circ b = d(e(f,b),b),$$

respectively. Exploiting the 2D interpretation once again, we come to the conclusion that the opening operation yields the locus of points covered by all instances of the structuring element b lying completely inside f, while the result of the closing operation is the complement of opening the complement of f by the structuring element b reflected through its center (in our case this is the coordinate system zero). The last operations to be defined are the morphological bottom-hat and top-hat transforms given by

(5) $$B_b(f) = f - f \bullet b,$$

(6) $$T_b(f) = f - f \circ b,$$

respectively. The above operations can be used to suppress shot noise which is not formed by single, isolated outliers, but rather spikes of positive width due to frequent sampling. It can be shown that these transforms extract those peaks and dips in the graph of f which cannot contain the shape of the graph of the structuring element b [7]. Hence, the shape of b determines the shape of detected spikes. In the proposed algorithm only the top-hat transform will be used.

Principles of singular spectrum analysis for 1D signals

Let f denote the same signal as before, let H be the window size in which f is analyzed and let $M<H$ be the range of the autocorrelation function. In this section it is assumed that the length of the signal f is equal to H. When the signal is longer, only H most recent samples are taken into consideration, and when it is shorter, all missing samples are assumed to be zero. Let us define the trajectory matrix F to be the $M\times(H\text{-}M\text{+}1)$ matrix whose entries are given by the formula

$$F_{kl} = f(k - l + H - 2M + 1)\,. \tag{7}$$

This matrix is Toeplitz and its rows are formed by lagged f signals [8]. Namely, the top row of this matrix are the last H-M+1 samples, the second row are the last but one H-M+1 values, and so on. The central object of interest in the singular spectrum analysis is the correlation matrix FF^T of rows of F, where the superscript ^{T} denotes the matrix transposition. This matrix is symmetric and positive-semidefinite, hence all its eigenvalues λ_i are real and non-negative, and all its eigenvectors v_i are also real and pair wise orthogonal. Without loss of generality it can be assumed that these eigenvalues are numbered in the descending order $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_M$ and that the eigenvectors are normalized to unity with respect to the Euclidean norm. The main idea of the singular spectrum analysis is to analyze the principal components p_i of F defined as [9]

$$p_i = v_i v_i^T F\,. \tag{8}$$

In this sense the singular spectrum analysis is similar to the well-known principal component analysis. Due to completeness of the eigenbasis $\{v_i\}_{i=1,\ldots,M}$, the sum of all principal components is equal to F, and it can be shown that the trend is contained in the principal component corresponding to λ_i of largest magnitude [10]. The rest of components with higher numbers carry information on details of the signal. Here we assume that we obtain acceptable accuracy just by taking the first component p_0 only. Although F is Toeplitz, in general no principal component maintains this property. Hence, to reconstruct the trend from the p_0 component, we must restore its property of being Toeplitz. We accomplish this by averaging p_0 in the diagonal direction. Therefore, the last step of the trend extraction procedure is to compute the trend $P_0(f)$ itself using the following formula

$$P_0(f)(k - l + H - 2M + 1) = \frac{1}{M} \sum_{k-l+H=2M+1} (p_0)_{kl}\,. \tag{9}$$

To avoid boundary effects arising due to incomplete knowledge on signal's past and future waveform, the first and last M-1 samples of $P_0(f)$ should not be used. In practical applications it appears that even more values should be dropped. In the proposed algorithm $2M$-1 values are dropped and the trend is delayed with respect to the original, noisy signal by exactly $2M$-1 samples.

The algorithm for signal despiking and trend extraction

Having all the above definitions in mind we are ready to sketch an algorithm for efficient signal despiking and trend extraction. The algorithm is very simple and consists of the following two steps:

1. In the first step sharp peaks and dips are removed from the input signal using the tophat transform. It was noted that when peak removal is conducted as first, broader plateaus of relatively small amplitude may appear in the place of removed peaks. On the other hand, when dip removal is conducted as first, in the place of removed dips broader sinks may sometimes be found. To remedy this, both sequences are computed: at one time removal of peaks is followed by removal of dips, and the other time removal of dips is followed by removal of peaks. These two results are then averaged yielding the despiked signal.

2. In the second step the trend of the despiked signal is computed using Eq. (9). When the algorithm is used for streaming input, at each step only a single next value is produced putting k-l+H-$2M$+1=H-$2M$. This means that the trend is delayed with respect to the on-line signal by $2M$-1 samples.

The above two steps can be rewritten in terms of a sequence of mathematical formulas to be computed. The first step reads

$$f_{\wedge\vee} = f - T_b(f) + T_b(T_b(f) - f)\,, \tag{10}$$

$$f_{\vee\wedge} = f + T_b(-f) + T_b(T_b(-f) + f)\,, \tag{11}$$

$$f_{despiked} = \frac{f_{\wedge\vee} + f_{\vee\wedge}}{2}\,. \tag{12}$$

while the second step is simply

$$Trend(f) = P_0\left(f_{despiked}\right), \tag{13}$$

When started, the algorithm gathers $2M$ samples to avoid boundary effects and the first f of length H is constructed using zero-padding. At every next time instant the algorithm computes a new f using the new sample and H-1 recent samples from the old f and calculates $Trend(f)(H\text{-}2M)$ which is output as the next value of the trend. This simple procedure can be easily coded in MATLAB; the corresponding code is presented in Figs. 1, 2, and 3.

```
% Despiking and trend extraction
% It is assumed that f is already of length >= H
function extracted=trend(f,b,M,H)

% Trim f
if (length(f)>H), f=f(end-H+1:end); end
% Remove top spikes first
despiked_f1=f-tophat1D(f,b);
% Remove bottom spikes
despiked_f1=despiked_f1+tophat1D(-despiked_f1,b);
% Remove bottom spikes first
despiked_f2=f+tophat1D(-f,b);
% Remove top spikes
despiked_f2=despiked_f2-tophat1D(despiked_f2,b);
% Average despiking results
despiked_f=(despiked_f1+despiked_f2)/2;
% Compute the SSA approximation
extracted=P0(despiked_f,M);
end
```

Fig. 1. Unoptimized MATLAB code for $Trend(f)$ given by Eq. (13). It is presumed that f is of length $\geq H$.

```
% f – the input signal, M – autocorrelation range
% function single_trend_value=P0(f,M)

% The matrix F is formed
H=length(f);
K=H-M+1;
F=zeros(M,K);
for i=1:M
    F(i,:)=f(M-i+1:H-i+1);
end
% SVD decomposition
[v,lambda]=eig(F*F.');
% Sorting eigenvalues and eigenvectors
[lambda,i]=sort(-diag(lambda));
v=v(:,i);
% Reconstruction of the next value of the trend
p0=v(:,1)*v(:,1).'*F;
single_trend_value=0;
for m=1:min(M,K)
 next_val=next_val+p0(m,K-M-m+1); % H>=2M
end
next_val=next_val/min(M,K);
end
```

Fig. 2. Unoptimized MATLAB code for $P_0(f)$ given by Eq. (9). It is presumed that f is of length H.

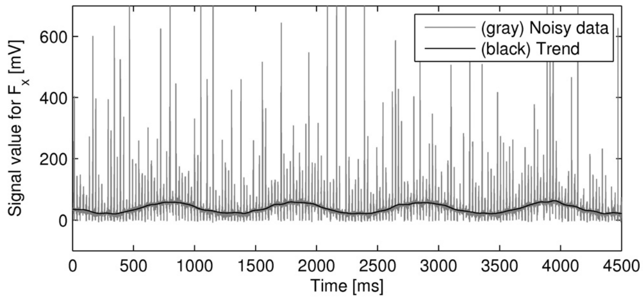
```
% Dilation of f by b of length 2a+1 and symmetric
function dilated=dilation1D(f,b)

dilated=zeros(size(f));
a=(length(b)-1)/2; l=length(f);
for k=1:length(f)
    f_chunk=f(max(k-a,1):min(k+a,l));
    b_chunk=b;
    if k-a<1, b_chunk =b(1-k+a+1:end); end
    if k+a>l, b_chunk =b(1:l-k+a+1).'; end
    dilated(k)=max(f_chunk+ b_chunk);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Erosion of f by b of length 2a+1 and symmetric
function eroded=erosion1D(f,b)

eroded=zeros(size(f));
a=(length(b)-1)/2; l=length(f);
for k=1:l
    f_chunk=f(max(k-a,1):min(k+a,l));
    b_chunk=b;
    if k-a<1, b_chunk=b(1-k+a+1:end); end
    if k+a>l, b_chunk=b(1:l-k+a+1); end
    eroded(k)=min(f_chunk-b_chunk);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Closing of f by b of length 2a+1 and symmetric
function closed=closing1D(f,b)

closed=erosion1D(dilation1D(f,b),b);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Opening of f by b of length 2a+1 and symmetric
function opened=opening1D(f,b)

opened=dilation1D(erosion1D(f,b),b);
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Top-hat of f by b of length 2a+1 and symmetric
function tophatted=tophat1D(f,b)

tophatted=f-opening1D(f,b);
end
```

Fig. 2. Unoptimized MATLAB code for morphological operations given by Eqs. (1)-(6).

The results

The proposed algorithm was applied to tensometric voltage measurements during material cutting. The measured voltage signals represent two orthogonal components of the force (referred to as F_x and F_z) acting on material being cut. The tensometer was glued in a close vicinity of plasma cutting area and the measurements were taken during operation of a 30kW plasma torch. It means that the observed values are substantially disturbed by strong electromagnetic fields of wide spectrum acting on the piezoelectric active element and electronics of the tensometer. More, the system was subject to unpredictable mechanical shocks which resulted in constant presence of parasitic resonances additionally polluting the output signal. In effect, the raw signal contains shot noise with SNR exceeding the value of 10 and stochastic noise of much smaller amplitude. Hence, this signal satisfies conditions a) and b) set in the introduction and can be treated by the proposed algorithm.

Fig. 4 below shows typical measured signals and results of the trend extraction algorithm. From Hooke's law it follows that only positive values of the signal are physically possible. However, due to noise also negative values were observed. The F_x signal was thresholded at zero for analysis, because useful component of this signal is well separated from zero.



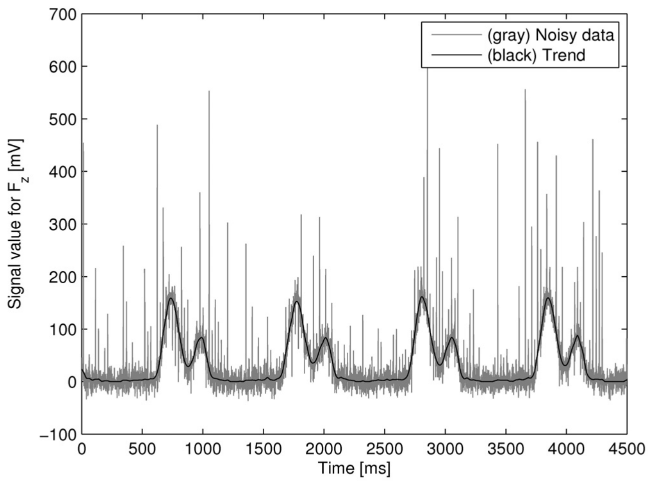


Fig. 4. Measured noisy signals (gray) and trends extracted using the proposed algorithm (black). The upper (lower) plot depicts the case of the signal to be interpreted as the x-component (z-component) of the force. In both cases the structuring element for morphological despiking was b=[0,15,30,15,0]. Parameters for trend extraction by means of the singular spectrum analysis was H=960 for both plots, and M=64 for the upper plot and M=24 for the lower plot.

Fig. 5 below shows 100-samples-long snapshots of raw signals, results of morphological despiking given by Eq. (12), and the trend given by Eq. (13). It is clearly seen that as far the raw signal does not contain spikes of amplitude greater than the pre-set value of thirty, the proposed morphological despiking scheme leaves it in untouched form; spikes are removed selectively. The extracted trend is therefore adequate and not affected by outliers.

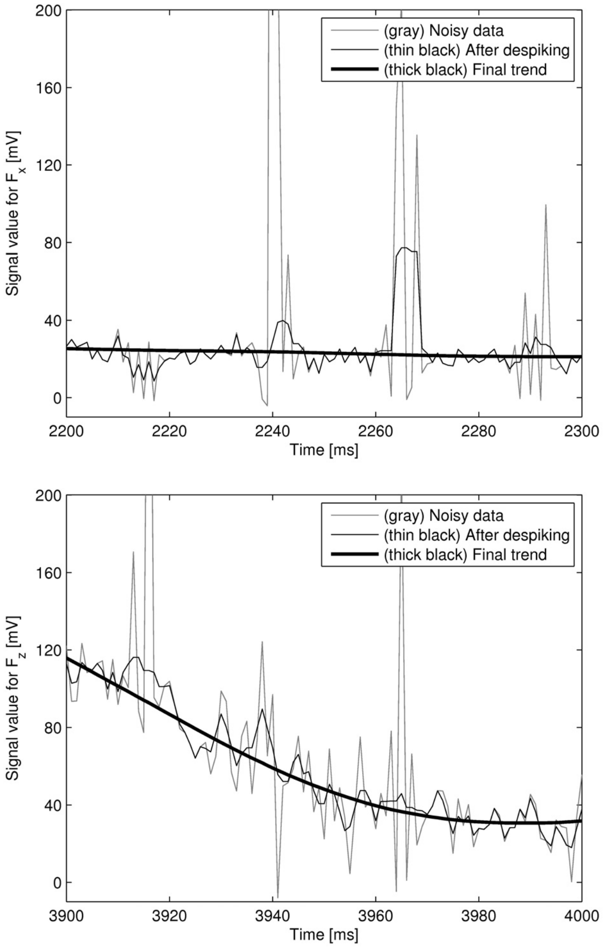


Fig. 4. Short snapshots of measured noisy signals (gray), despiking results (thin black), and trends extracted using the proposed algorithm (thick black).

Conclusions

We proposed an algorithm for effective, real-time denoising of signals containing two types of noise: shot noise of large amplitude and non-trivial waveform and non-stationary stochastic noise of zero mean. The algorithm was tested on signals measured in a system subject to strong electromagnetic and mechanical disturbances. The obtained results prove that the algorithm is robust and can be easily adapted to various signal characteristics just by adjustment of its parameters; it can be treated as an alternative to the empirical mode decomposition method which is often used in presence of severe noise [11]. Considering conceptual and coding simplicity of the algorithm, it can be easily modified to treat also other classes of discrete signals. A drawback of relative computational complexity seems to pose no problem for current microprocessors and CPUs.

Acknowledgments

The study was supported by the Proteus Project co-financed by the European Regional Development Fund within the framework of the 1. priority axis of the Innovative Economy Operational Programme, 2007-2013, submeasure 1.1.2 "Strategic R&D Research" (contract no. POIG.01.02.01-00-014/08).

REFERENCES

[1] Haykin S., Unsupervised adaptive filtering. Vol I: Blind source separation, *Wiley*, 2000

[2] To A., Moore J., Glaser S., Wavelet denoising techniques with applications to experimental geophysical data, *Sign. Proc.*, 89 (2009), No. 2, 144-160

[3] Kim J., Lyon R., Cepstral analysis as a tool for robust processing, deverberation and detection of transients, *Mech. Syst. Sign. Proc.*, 6 (1992), No. 1, 1-15

[4] Bustos O., Yohai V., Robust estimates for ARMA models, *J. Am. Stat. Assoc.*, 81 (1986), No. 393, 155-168

[5] Nobre F., et al., Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology, *Stat. Med.*, 20 (2001), No. 20, 3051-3069

[6] Chanda B., Morphological algorithms for image processing, *IETE Tech. Review*, 25 (2008), No. 1, 9-18

[7] Soille P., Morphological image analysis. Principles and applications, *Springer*, 2003

[8] Wei W., Time series analysis. Univariate and multivariate methods, *Addison-Wesley*, 2005

[9] Hassani H., Singular spectrum analysis: methodology and comparison, *J. Data Sci.*, 5 (2007), 239-257

[10] Allen M., Smith L., Optimal filtering in singular spectrum analysis, *Phys. Let. A*, 234 (1997), No. 6, 419-428

[11] Jakubowski J., Michalski A., Metoda dekompozycji na mody empiryczne w eliminacji szumów prądu cewki przepływomierza elektromagnetycznego dla kanału otwartego, Przegląd Elektrotechn., 11 (2009), 111-114

***Author**: MSc Konrad Bojar, Industrial Research Institute for Automation and Measurements, Laboratory of Autonomous Defense Systems, al. Jerozolimskie 202, 02-486 Warsaw, Poland, e-mail: kbojar@piap.pl.*