Blocked and Accelerated Wavelet De-noising Algorithm Based on Data Splitting and Wavelet Analysis in Large Data Environment for Aero-Engine Health Monitoring

Chuanchao Zhang
School of Information Engineering, Wuhan University of Technology, Wuhan, PR. China
Aviation Industry Corporation of China, Beijing, PR. China
Email: 574799770@qq.com

***Abstract*—Data de-noising is a necessary part of health management, and it is the premise and foundation of effective feature extraction, condition monitoring and fault diagnosis for aero-engine. Random noise can cause serious interference to effective signals, and even lead to signal distortion and misdiagnosis of health condition. In view of the contradiction between the limited computing power of aircraft airborne system and the large amount of data processing, an blocked wavelet de-noising algorithm for large data is proposed based on the principle of data splitting theory and the wavelet theory under the multiple constraints of large data, high de-noising precision and fast processing speed. The algorithm used data splitting principle to split large data into small data sets, reduced the computational requirements of large data, and accelerated the speed of wavelet de-noising. The processing results of the theoretical data and the actual airborne aero-engine monitoring data showed that, compared with the traditional algorithms, the algorithm can protect the effective information and maintain the same de-noising accuracy, and the data de-noising time in the aero engine health monitoring data environment was accelerated by 4 times at least.**

***Index Terms*—aero-engine, health management, large data, data splitting, wavelet theory, modulus maxima, random noise, SNR**

I. Introduction

In the process of data acquisition and transmission, because of the interference of a variety of factors, the collected data often contain noise, it is necessary to remove the noise, so as to facilitate subsequent process and analyze. As a most important prerequisite for accurate data reconstruction, data de-noising is an important research topic.

For a long time, scholars have studied data de-noising from space to time transform domain and non local and so on, and put forward many effective solutions. They are space domain de-noising methods such as neighborhood filtering [1], [2] and Partial Difference Equation (PDE) [3], [4], and transform domain de-noising methods such as wavelet transform [5]-[7] and multiscale analysis [8], [9], and non local de-noising technology [10]-[14] such as Non-local Means (NLM) and Principle Neighborhood Dictionary (PND) [15] and Local Pixel Grouping-principal Component Analysis (LPG-PCA) [16] and Block-Mactching and 3D filtering (BM3D) [17]. These methods are effective to de-noise for small data.

In recent years, signal sparse representation has become a hot topic in the field of signal processing. The redundant sparse de-noising methods are represented by K-Singular Value Decomposition (K-SVD) [18]-[24] and Multiscale K-SVD (MK-SVD) [25], [26] and Learned Simultaneous Sparse Coding (LSSC) [27]. They can satisfy with the sparsity, feature retention and separability of signal noise, and has been successfully applied to the image de-noising [28], [29] and seismic data de-noising [30]. As a new method, redundant sparse de-noising method is used to sparse the de-noise signal with redundant atom library to achieve the purpose of removing noise. It has shown great potential in the image de-noising [31]-[34] and seismic data de-noising [35], [36]. While in the traditional sequential sequence of massive data processing, it is difficult to make a breakthrough under time constraint while ensuring the de-noising accuracy.

The wavelet transform has the following advantages: (1) Wavelet decomposition can cover the whole frequency domain; (2) Wavelet transform can greatly reduce or remove the correlation between different features by selecting appropriate filters; (3) Wavelet transform has the characteristic of "zoom", which can be used in high frequency resolution and low time resolution in low frequency sections (wide analysis window). Low frequency resolution and high time resolution (narrow analysis window) can be used in the high frequency sections; (4) The wavelet transform can be implemented by fast algorithms, such as Mallat Wavelet Decomposition Algorithm [37]. Therefore, wavelet analysis has been applied to data de-noising effectively [37]. When the noise is de-noising, the wavelet analysis

Manuscript received August 18, 2018; revised December 5, 2018.

doi: 10.18178/ijoee.6.2.79-87

theory is not only the precision is guaranteed, but also processing speed is relatively fast. However, when dealing with the large data for aero-engine health management, it can't meet the requirements of the data processing under the conditions of large data, high de-noising precision, fast processing speed and so on.

In view of the advantages of wavelet analysis theory, this paper uses the principle of data splitting and proposes a blocked and accelerated wavelet large data de-noising algorithm based on data splitting and wavelet analysis (called as BAWLDDABDPW). It can divide large data sets into small data sets and accelerate the wavelet data de-noising algorithm, and solve the data processing requirements under the multiple constraint conditions such as large data, high de-noising precision, fast processing speed and so on.

II. Wavelet Theory

Signal and noise have different characteristics on different scales. On the base of this principle, Mallat, XU and Donoho have proposed respectively a signal filtering algorithm based wavelet. At present, wavelet filtering methods are mainly divided into Bayesian method and non Bayesian method. In which the non Bayesian method is divided into three kinds: Modulus maxima reconstruction filter proposed by Mallat [38]; Wavelet domain threshold filtering proposed by Donoho [39]; Spatial correlation filter, that is scale space correlation filtering, proposed by Xu [40]. The de-noising method based on wavelet analysis has a good effect in data de-noising [41]-[45], including modulus maxima de-noising [46]-[48], threshold de-noising [49]-[51] and correlation de-noising method [40].

A. Wavelet Transform Theory

If $\psi(t)$ is a square integrable function, that is $\psi(t) \in L^2(R)$. If its Fourier transform satisfies the condition:

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \tag{1}$$

where: $\Psi(\omega)$ is $\psi(t)$'s Fourier transform. $\psi(t)$ is called a basic wavelet or a wavelet generating function. Equation (1) is called as the admissible condition of a wavelet function, and the generating wavelet is not unique, but optional.

If Signal $f(t) \in L^2(R)$, then its continuous wavelet transform is defined as:

$$WT_f(a,\tau) = \left[f(t), \psi_{a,\tau}(t) \right] = \frac{1}{\sqrt{a}} \int_R \bar{\psi}\left(\frac{t-\tau}{a}\right) dt \tag{2}$$

If $\psi_{a,\tau}(t)$ satisfies admissible condition equation (1), then continuous wavelet has inverse transformation, its corresponding inverse transformation equation is:

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} (W_\psi f)(a,\tau) \left[\frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) \right] db \right\} \frac{da}{a^2} \tag{3}$$

In actual signal processing, it is necessary to carry out discrete wavelet transform, that is $a = a_0^0, a_0^1, \cdots, a_0^j, j = 0,1,2,\cdots, N-1$. Normally, if $a = 2^J, j = 0$, then the signal is sampling frequency. If $j = 1$, the requency is divided by two. And the like.

B. De-noising Algorithm Based on Modulus Maxima of Wavelet Transform

The principle of modulus maximum de-noising is to de-noise based on the difference between the truth value and the singularity characteristic of noise. Singularity and singular point are important features of signals. Wavelet transform has the ability to determine the location and size of signal singularity. The description of singularity is related to the Lipschitz index.

【Definition 1】 (Lipschitz index) Function $f(t)$ owned pointwise (or local) Lipschitz index $\alpha(\alpha \ge 0)$ in point t_0. If there are $K > 0$ and an polynomial $P_{t_0}(t)$ which is powered by $n = [\alpha]$, they satisfy

$$\left| f(t) - P_{t_0}(t) \right| < K |t - t_0|^\alpha, \forall t \in R \tag{4}$$

If function $f(t)$ owned pointwise Lipschitz index α for $t_0 \in [a,b]$, where K and t_0 are irrelevant, then function $f(t)$ owned coherent Lipschitz index $\alpha(\alpha \ge 0)$ in the interval $[a,b]$.

Notice: Lipschitz is an expansion for continuity depiction. If $f(t)$ is continuous and differentiable with power $n = [\alpha]$ in a neighbourhood of t_0, then $P_{t_0}(t)$ is a polynomials $n+1$ ahead of the Taylor expansion in t_0. That is:

$$P_{t_0}(t) = \sum_{k=0}^{n} \frac{f^{(k)}(t_0)}{k!} (t - t_0)^k \tag{5}$$

where, $f^{(k)}(t_0)$ is the k-order derivative of $f(t)$ in t_0, $k = 0,1,2,\cdots,n$.

Specially, if function $f(t)$ is continuous and differentiable in t_0, then the Lipschitz index of $f(t)$ in the point is 1. If $f(t)$ is differentiable in t_0, but the derivative of $f(t)$ is bounded and discontinuous, then the Lipschitz index of $f(t)$ in the point is 1 still. If $f(t)$ is bounded and discontinuous in t_0, then the Lipschitz index of $f(t)$ in the point is 0.

If the Lipschitz index of $f(t)$ less than 1 in the point, then function $f(t)$ is singular in t_0 point. If the Lipschitz index α of $f(t)$ in the point t_0 satisfies $n < \alpha < n+1$, then $f(t)$ is a n differentiable function, but the n order derivative $f^{(n)}(t_0)$ is singular in t_0 point, its Lipschitz index is $\alpha - n$. This can describe the singularity of the signal. The range of Lipschitz index can expand to $-1 \le \alpha < 0$. If the schutz index for Primitive Function $F(t)$ of $f(t)$ in t_0 is $\alpha + 1$ ($-1 \le \alpha < 0$), then Lipschitz index of $f(t)$ in t_0 is α. Negative Lipschitz index means that the function has greater singularity than discontinuous function ($\alpha = 0$). For example:

- A measurement of singular point for polygonal function. If t_0 is singular point, then Lipschitz index in t_0 point is $\alpha = 1$.
- A measurement of singular point for jump function. If t_0 is step point, then Lipschitz index in t_0 point is $\alpha = 0$.
- A measurement of function singularity. If $f(t)$ is a singular function, then its Lipschitz index in t_0 =0 point is $\alpha = -1$.

S. Mallat combines the local singularity of the function with the modulus local maxima of the wavelet transform.

【Theorem 1】If wavelet $\psi(t)$ are real functions and continuous, with attenuation

$$\psi(t) \le K\left(1+|t|\right)^{-2-\varepsilon}, \varepsilon > 0 \tag{6}$$

and $f(t) \in L^2(R)$ on the interval I owns unanimous Lipschitz index α, $-\varepsilon \le \alpha \le 1$, then there is a constant $C > 0$ satisfying $\forall a, b \in \mathrm{I}$. There is,

$$\left|Wf(a,b)\right| \le Ca^{\alpha+\frac{1}{2}} \tag{7}$$

Conversely, for a certain α, $-\varepsilon \le \alpha \le 1$, if the wavelet transform of $f(t) \in L^2(R)$ satisfies the above formula, then $f(t)$ on the interval I owns unanimous Lipschitz index α.

If t_0 is the singular point of $f(t)$, then $\left|Wf(a,b)\right|$ can take maximum value in the point $b = t_0$. Different singular points can be found at different scales. The singularities of noise and signal are hidden in these singular points, and they vary with different scales.

【Theorem 2】 $f(t)$ owns unanimous Lipschitz index α on the interval I, for $\alpha = 2^j$, it is obtained as follows:

$$\left|Wf(2^j, b)\right| \le C2^{j\left(\alpha+\frac{1}{2}\right)} \tag{8}$$

Then, when $\alpha > -\frac{1}{2}$, modulus maxima of wavelet transform is increasing with the increasing of scale j. When $\alpha < -\frac{1}{2}$, modulus maxima of wavelet transform is decreasing with the increasing of scale j. When $\alpha < -\frac{1}{2}$, this indicates that the signal is more singular than the discontinuous signal, which is the corresponding situation of noise, such as white noise $\alpha = -\frac{1}{2} - \varepsilon,\ \forall \varepsilon > 0$. Therefore, the modulus maxima of wavelet transform can be used to distinguish signals and noises as the changing scales.

The modulus maxima de-noising method is, based on the wavelet transform of the observation function $f(t)$, and the different trends for singular points of real value $x(t)$ and noise $e(t)$ vary with the different scales, so that the singular points are removed coming from the de-noising, and then the singular points corresponding the real values are reconstructed to achieve the de-noising purpose.

III. Data Splitting

The so-called data set splitting is to divide a large scale data set into several smaller disjoint subsets by a certain splitting rule in the case that the classification effect is not affected as much as possible. Assuming that the original dataset is S, the subset of each data is S_i, then

$$S = S_1 + S_2 + \cdots + S_n = \sum_1^n S_i \tag{9}$$

where, $S_i \cap S_j = \phi(i \ne j)$.

Now, the splitting rules includes random splitting method [52], sequential splitting method [52], attribute value splitting method [52], rough set theory splitting method [52], information entropy splitting method [53], mean shift algorithm [54], etc. The random splitting method consists of a small set of data randomly selected from a large data set and repeats the process until all the data are splitted. Sequential splitting method is extracting the former m (m<n) data in order to form a small data set according to priority from a large data set with N data, and repeating the process until all the data in the original dataset is split out. The attribute value splitting method splits the original big dataset according to the values of a certain attribute or groups of attributes. For example, if an attribute in a dataset has a value range of 1 to 10, the dataset can be splitted into ten subsets according to each value of the attribute. Rough set theory splitting method is to make a preliminary calculation of the dataset before splitting, and to get a reduction of the set, and then calculate the importance of the attribute of the reduction set, and to sort according to the importance of the attribute, and to classify the data and to split dataset on this basis, to ensure the same knowledge or rules between

the information system after the splitting and the original system, to efficiently solve the splitting of high-dimensional massive data. In view of informatics, information entropy splitting method uses information entropy theory, to select the splitting method which makes the total entropy of data subset smaller. Mean shift algorithm is a two clustering and processing splitting method. Firstly, it is used to pre-splitting the large scale image dataset, and then the hierarchical clustering algorithm is used to cluster the pre-splitting image again. When the mean shift algorithm is used for image splitting, according to the result of the image smoothing, all the pixels converged at the same density maximum point are used as the same class, and give the same indicia to used pixel points in the class. If the number of points for a class is less than the minimum value M, the class is removed.

IV. Algorithm Research

A. Random Noise Suppression Model

The classical data de-noising principle is described as follows: The original data $f(t)$ is contaminated by additive Gauss white noise with zero mean deviation of σ. If $f(t)$ is the collected noise data, then

$$f(t)=x(t)+\varepsilon(t) \tag{10}$$

The goal of de-noising is to design a de-noising method to remove the noise $e(t)$ in $f(t)$, so that the de-noised data is as close as possible to the original data $x(t)$.

At present, the non local de-noising method has achieved good effect on the removal of mild and small noise (small σ) in the data, but it is still a challenge to recovery the data polluted by the serious noise (large σ). When the noise is low, the data de-noising faces three prominent contradictions: the first, the data features are kept as possible as. The second, the pseudo structure which is generated from the noise should be avoided as far as possible, thus the artificial noise is reduced. The third, how to deal with the de-noising problem of large data.

B. Algorithmic Description

1) Algorithm procedures

The basic steps of the blocked modulus maximum data de-noising algorithm based on data splitting and wavelet analysis are as follows:

The first step, for the noisy raw dataset S is splitted into subsets $S_i\left(i=1,2,\cdots,n\right)$ by using a data splitting method.

The second step, each data subset is to de-noise by wavelet de-noising based on modulus maximum de-noising method. The specific steps are described as follows:

① The wavelet transform of noisy signals is carried out, the scale is $S=2^j, j=1,2,\cdots,J$, and the modulus maxima of transformation coefficients at each scale are obtained.

② Starting from the maximum scale (for example T), a threshold T is set. If the modulus maximum of this scale less than T, then the modulus maxima point is removed and the other is retained. And then, a new set of modulus maxima points on the maximum scale is obtained.

③ A neighborhood of each maximum point retained in a scale function $j=J$ is taken, for example $N\left(t_i,\varepsilon_J\right)$. The corresponding to the maximum points in the neighborhood $N\left(t_i,\varepsilon_J\right)$ on the scale $j-1$ is found, and these maximum points are retained and other maximum points are removed. Thus, a set of new maximum points on the scale $j-1$ is obtained.

④ To set $j=j-1$ and to repeat the step ③ until $j=2$.

⑤ On the maximum value points are preserved when $j=2$, the corresponding maximum points are found out, and the others are removed.

⑥ The wavelet coefficients of retained maximum points at multiple scales are reconstructed by appropriate methods. For example, signal reconstruction methods include the cross projection method proposed by Mallat [55], and the fast algorithm of approximate signal reconstruction using frame theory [37].

⑦ The reconstructed data subsets SW_i are obtained.

The third step, the result SW_i of each subset after de-noising are reconstructed, and the whole dataset SW after de-noising is obtained, that is $SW=\bigcup_{i=1}^{n} SW_i$.

In discrete wavelet transform, $Wf\left(j,k\right)$ can be noted as $W_{2^j}f[k]$. If $k=m$ is a modulus maximum point, then

$$\begin{cases}\left|W_{2^j}f[m]\right|\geq\left|W_{2^j}f[m-1]\right|\\ \left|W_{2^j}f[m]\right|\geq\left|W_{2^j}f[m+1]\right|\end{cases} \tag{11}$$

And the equal signs can’t be taken at the same time in the above two formulas.

2) Algorithmic description

The blocked modulus maximum data de-noising algorithm based on data splitting and wavelet analysis are described as follows as Table I:

TABLE I. BLOCKED MODULUS MAXIMUM DATA DE-NOISING ALGORITHM BASED ON DATA SPLITTING AND WAVELET ANALYSIS

Input: noise dataset S, numbers of data subset n, the size of data subset N output: SW after de-noising
1 To take data splitting, $S = S_1 + S_2 + \cdots + S_n = \sum_1^n S_i$. 2 *for* S_i, do ①To take Wavelet transform for noise signal, the scale is $S = 2^j, j = 1, 2, \cdots, J$. ②To obtain a new set of modulus maxima points on the maximum scale. ③To use $N(t_i, \varepsilon_J)$ when $j = J$ and obtain the modulus maxima points when $j = j-1$. ④To set $j = j-1$, and repeat the step ③, and until $j = 2$. ⑤To use $N(t_i, \varepsilon_J)$ when $j = 2$ and obtain the modulus maxima points when $j = 1$. ⑥To reconstruct wavelet coefficients. ⑦To output reconstructed SW_i data subsets. *while* $i \le n$ 3 To reconstruct the whole dataset SW after de-noising, that is $SW = \bigcup_{i=1}^{n} SW_i$.

V. EXPERIMENTAL ANALYSIS AND RESULTS

A. *Experimental Data and Environmental Conditions*

1) Experimental data

Aero-engine is the most critical and complex part of aircraft. Its performance and safety will directly affect the performance and safety of the whole aircraft. In order to improve the safety and reliability of the aircraft and its aero-engine, the advanced condition monitoring technology is used to measure the health condition of the aircraft and aero-engine during the running of the aero-engine. The fault signs are detected in the early time, and the fault diagnosis and maintenance are carried out in advance to ensure that the fault is eliminated in time.

The measured data in one flight of a certain aero-engine is selected to carry out simulation and verification. The flight continued nearly 13h, and the data sampled points are 46595, and 46080 of sampling data are selected. There are 205 parameters for monitoring aero-engine, and the sampled data occupies 116MB (121,864,192 bytes) space. Generally, the measurement sample of Static Temperature (SAT) parameter is chosen as the analysis object.

2) Simulation environment

A series of simulation experiments are carried out by using the de-noising algorithm. The hardware test platform is Intel Core i7 CPU with 4.0GHz main frequency and 16GB memory. The simulation software platform is Windows 7 64bits operating system and Matlab 2014a.

3) Data splitting rules and parameters setting

Data subset size selection is based on aero-engine data transmission characteristics and wavelet analysis scale requirements. The aero-engine monitoring system has transmitted downwards at one frame per time, with 60 sampling points per frame. And the scale requirements of wavelet transform meet the requirements $S = 2^j (j = 1, 2, \cdots, J)$ and $S > 1024$. Therefore, the smallest size of the data subset is the least common multiple of them, which is 3840. The original dataset is divided into 12 subsets, which is $n=12$.

It is assumed that the noise is standard zero mean Gauss white noise and noise intensity $\sigma \in \{5, 10, 15, 20, 25, 50, 100\}$.

B. *Performance Evaluation Index*

According to the requirements of aero-engine health management, the following 4 parameters are used to measure the performance of the algorithm.

1) Signal to Noise Ratio (SNR)

SNR is used as an objective evaluation criterion for de-noising performance. The definition of SNR [20] is as follows:

$$SNR = 10\lg\left(\frac{\sum_i x_i^2}{\sum_i (x_i - y_i)^2}\right) \tag{12}$$

where, x_i is the original signal which is N in length, and y_i is the signal after de-noising. The SNR value is the greater means that the signal de-noising effect is the better.

2) Mean Square Error (MSE)

The approximation degree of the two signals is evaluated by MSE. MSE is defined as follows:

$$MSE = \frac{1}{N}\left(\sum_{i=1}^{N}(x_i - y_i)^2\right) \tag{13}$$

3) Normalized Correlation (NC)

The degree of approximation of the two signals is evaluated by NC. The definition of NC is as follows:

$$NC = \frac{\sum_{i=1}^{N} x_i \bullet y_i}{\sum_{i=1}^{N}|x_i|^2 \bullet \sum_{i=1}^{N}|y_i|^2} \tag{14}$$

The NC value is closer to 1, the two signals are more similar.

4) Algorithm Running Time TIME

TIME is an important parameter to evaluate the performance of the algorithm. Under the premise of ensuring the algorithm accuracy, the shorter TIME of the algorithm is, the better performance of the algorithm is.

C. The Effect of Data Size on the Running Time of De-noising Algorithm

In order to study the effect of data size on the processing time of wavelet de-noising algorithm, the author selected 3840, 7680, 11520, 23040, 34560, 46080 aero-engine condition health test samples, that is, the data sizes are 1, 2, 3, 6, 9 and 12 times. Under the condition of the noise intensity σ =5, the processing time and accuracy of the algorithm are simulated. The simulation results are shown in Table II, Fig. 1 and Fig. 2.

TABLE II. PERFORMANCE COMPARISON OF MODULUS MAXIMUM DE-NOISING METHODS BEFORE AND AFTER DATA SPLITTING IN DIFFERENT DATA SIZES

data sizes		3840	7680	11520	23040	34560	46080
Before data splitting	SNR	101.32	102.09	102.25	101.75	102.26	101.80
	MSE	2.5855	2.39	2.356	2.478	2.354	2.485
	NC	0.9995	0.9996	0.9996	0.9996	0.9996	0.9985
	TIME(s)	169	519	1034	3490	10196	29740
after data splitting	SNR	100.39	99.56	99.76	100.12	100.39	99.82
	MSE	2.84	3.08	3.03	2.92	2.84	3.01
	NC	0.894	0.998	0.9990	0.9994	0.9994	0.9993
	TIME(s)	28.66	102.77	224.26	862.86	1919.80	3444.71

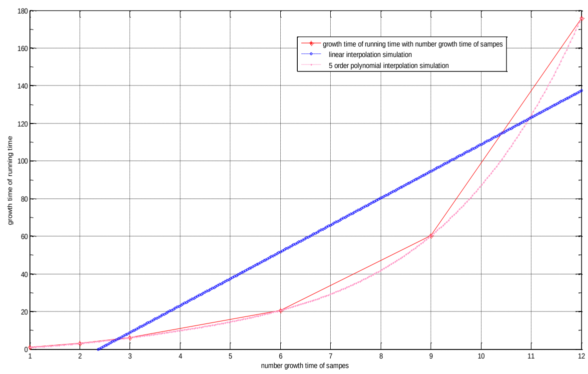


Figure 1. Effect of data scale growth on processing time of modulus maximum de-noising algorithm.

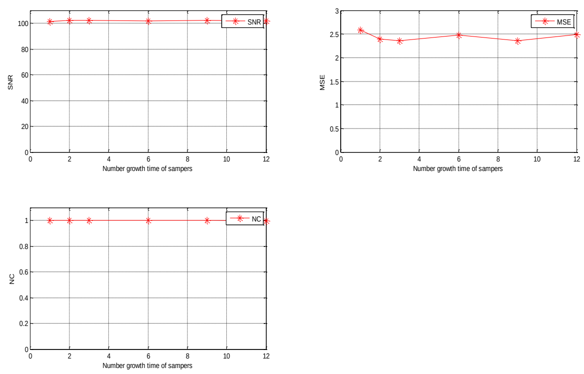


Figure 2. Effect of data scale growth on the performance of modulus maximum de-noising algorithm.

Under the premise of ensuring the fitting accuracy, the polynomial fitting for effect of data scale growth on processing time of modulus maximum de-noising algorithm is carried out. The fitting curve is as follows:

$$y = -6.7003e^{-06} \times x^5 + 0.018041 \times x^4 - 0.22364 \times x^3 + 1.3796 \times x^2 - 0.77262 \times x + 0.59866 \tag{15}$$

The linear fitting curve is as follows:

$$y = 14.276 \times x - 33.992 \tag{16}$$

The accuracy reaching 5.153e-14 of polynomial fitting is the highest according to equation (15), while the accuracy reaching 64.281 of linear fitting is the lowest according to equation (16). Thus, with the increasing of data size, the processing time of the wavelet de-noising algorithm is increasing by a power function with 5 constant.

D. The Effect of Noise Intensity on Running Time of De-noising Algorithm

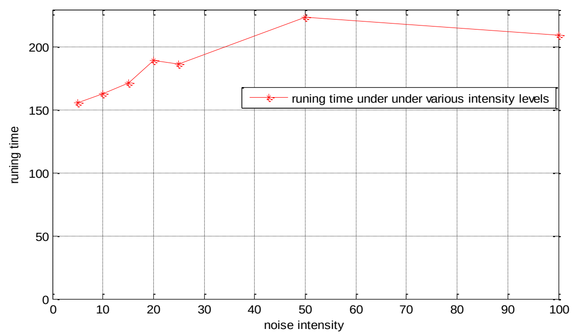


Figure 3. Time performance change of modulus maximum de-noising algorithm under different noise intensities.

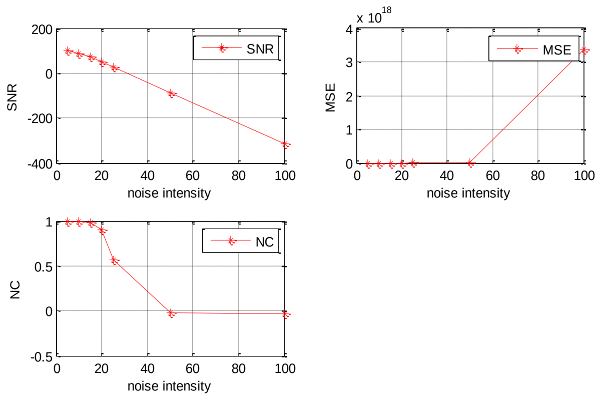


Figure 4. Accuracy performance change of modulus maximum de-noising algorithm under different noise intensities.

In order to analyze effects of the noise intensity of aero-engine on performance of the wavelet de-noising algorithm, 4096 test samples from the original noise data of the aero-engine are selected and analyzed. Generally, the noise type is Gauss white noise, and the standard intensity deviation is $\sigma \in \{5, 10, 15, 20, 25, 50, 100\}$, and the mean value of noise intensity is zero. Under the sets of noise intensity, the performance of the modulus maximum de-noising algorithm is simulated and shown in Fig. 3 and Fig. 4.

E. Acceleration Analysis of De-noising Algorithm by Data Splitting

Because the noise intensity has little effect on the accuracy and running time of the wavelet de-noising algorithm, it can be ignored. Therefore, the noise intensity can be fixed, and then the effect of data splitting on the performance and running time of the de-noising algorithm is analyzed. Generally, the noise intensity is set $\sigma \in \{5\}$ to simulate and compare the data before and after data splitting. Fig. 5 shows the effect of data splitting on the run time of wavelet de-noising algorithm. Fig. 6 shows the effect of data splitting on the accuracy of wavelet de-noising algorithm.

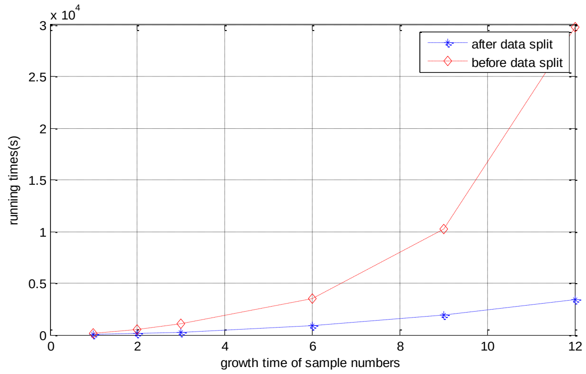
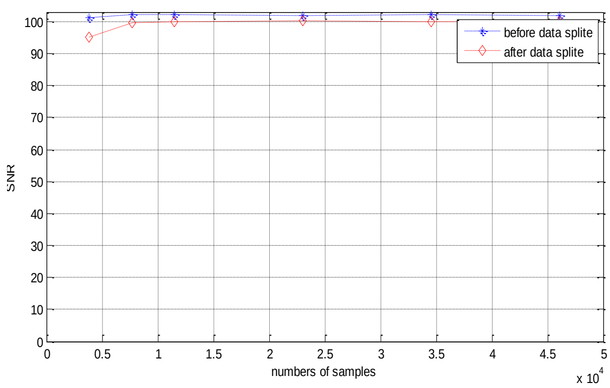
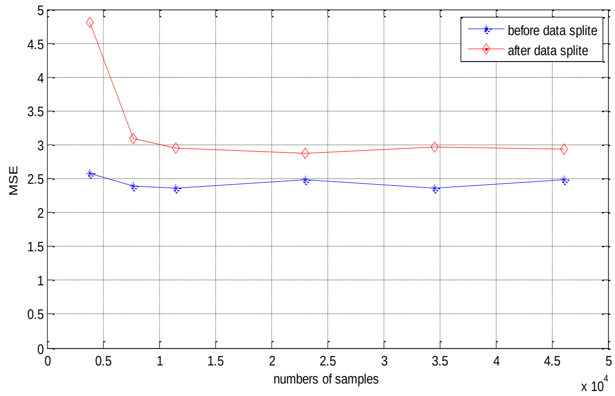


Figure 5. Effect of data splitting on run time performance of wavelet de-noising algorithm.

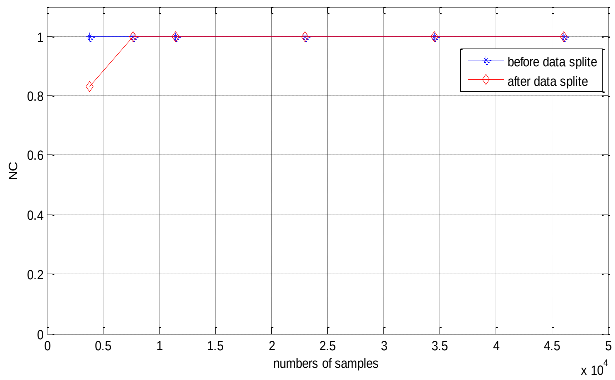
According to Fig. 5, with the increasing of data size, the processing time of data de-noising is increasing rapidly. After data splitting, the processing time of data de-noising has been effectively reduced. At the same time, according to Fig. 6, with the increasing of the size of the data, the SNR and MSE are reduced to some extent, while the NC remains basically the same. The reduction scope of SNR and MSE is not large and can be ignored. It can be explained that data splitting has no effect on the accuracy performance of the wavelet de-noising algorithm. According to Fig. 7, data splitting accelerates the processing time of data de-noising, and the speedup is over 4.



(a) SNR



(b) MSE



(c) NC

Figure 6. Effect of data splitting on de-noising accuracy performance of wavelet de-noising algorithm.

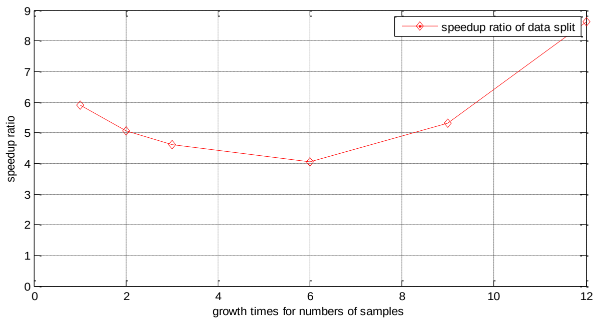


Figure 7. Acceleration effect of data splitting for wavelet de-noising algorithm.

VI. Conclusion and Prospect

The wavelet transform has a good time frequency localization and multi-resolution analysis ability. Due to the increasing of the wavelet transform modulus maximum of effective signal with the increasing of the scale, while the decreasing of the wavelet transform modulus maximum of the noise with the increasing of the scale, and the noise can be de-noising with the different features of the effective signal and the noise modulus maximum with the scale changing. This paper used the theory of data splitting and wavelet analysis, on the basis of a large dataset been splitted into small datasets, and used the wavelet transform based modulus maximum de-noising algorithm to de-noise the small data sets, and then reconstructed the data after de-noising. After simulation of real aero-engine monitoring data, not only the noise is well suppressed, but also the accuracy of de-noising is maintained, otherwise what is more important is that the following conclusions can be drawn under the large data environment of aero-engine monitoring:

(1) With the increasing of data size, the running time of the wavelet de-noising algorithm is increasing by a power function with 5 constant.

(2) Noise intensity has little effect on the accuracy and running time of the wavelet de-noising algorithm, it can be ignored.

(3) Data splitting accelerates the running time of data de-noising, and the speedup is over 4 at least.

The next step is to study the acceleration of data de-noising by sparse representation of large-scale signals.

Acknowledgment

The author wish to thank those who had studied data splitting and wavelet de-noising. This work was supported and inspired in part by a grant from them.

References

[1] A. V Oppenheim and R. W. Schafer, *Digital Signal Processing*, New York: Prentice Hall Inc., 1975.

[2] T. Sun and Y. Neuvo, "Detail-preservin median-based filters in image processing," *Pattern Recognition Letters*, vol. 15, no. 4, pp. 341-347, 1994.

[3] J. Yu and Y, Wang, "Image noise reduction based on anisotropic diffusion: A survey," *Journal of Electronic Measurement and Instrumentation*, vol. 25, no. 2, pp. 105-116, 2011.

[4] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear de-noising," *IEEE Trans. Image Process*, vol. 10, no. 2, pp. 231-241, 2001.

[5] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform Theory*, vol. 38, no. 2, pp. 617-643, 1992.

[6] D. L. Donoho, "Denoising by soft thresholding," *IEEE Trans. Inform Theory*, vol. 41, no. 3, pp. 613-627, 1995.

[7] J. L. Stack, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process*, vol. 11, no. 6, pp. 670-684, 2002.

[8] Y. Liu and W. Tang, "De-noising algorithm of AMSVD and its application in PQMD," *CEA*, vol. 52, no. 23, pp. 255-259, 2016.

[9] B. Song, L. Xu, and W. Sun, "Novel image transform method and its application in image de-noising," *Chinese Journal of Scientific Instrument*, vol. 29, no. 7, pp. 1455-1459, 2008.

[10] B. Cai, *et al.*, "An improved non-local means denoising algorithm," *PR&AI*, vol. 29, no. 1, pp. 1-10, 2016.

[11] Z. Huang, X. Fu, and W. Liu, "Non-local means denoising algorithm with hybrid similarity weight," *Journal of Computer Applications*, vol. 36, no. 2, pp. 556-562, 2016.

[12] Z. Liu, Z. He, and W. Guo, "A hybrid fault diagnosis method based on second generation wavelet de-noising and local mean decomposition for rotating machinery," *ISA Transactions*, pp. 1-10, 2015.

[13] S. Ning, Z. Han, and Z. Wang, "Application of sample entropy based LMD-TFPF de-noising algorithm for the gear transmission system," *Entropy*, vol. 414, no. 18, pp. 1-12, 2016.

[14] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 60-65.

[15] T. Tasdizen, "Principal neighborhood dictionaries for nonlocal means image de-noising," *IEEE Trans. Image Process*, vol. 18, no. 12, pp. 2649-2660, 2009.

[16] L. Zhang, W. Dong, and D. Zhang, "Two-stage image de-noising by principal component analysis with local pixel grouping," *Pattern Recognition*, no. 43, pp. 1531-1549, 2010.

[17] K. Dabov, A. Foi, and V. Katkovnik, "Image de-noising by sparse 3-d transform domain collaborative filtering," *IEEE Trans. Image Process*, vol. 16, no. 8, pp. 2080-2095, 2007.

[18] M. Elada and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process*, vol. 15, no. 1, pp. 3736-3745, 2006.

[19] G. Zheng, Q. Li, and G. Li, "Singular value decompositions signal de-noising algorithm based on genetic algorithm," *Application Research of Computer*, vol. 32, no. 8, pp. 2281-2285, 2015.

[20] Y. Wang, "Mean value of elgenvalue-based SVD signal de-noising algorithm," *Computer Applications and Software*, vol. 29, no. 5, pp. 121-123, 2012.

[21] F. Jiang, Z. Zhu, and W. Li, "Fault diagnosis of rotating machinery based on noise reduction using empirical mode decomposition and singular value decomposition," *Journal of Vibroengineering*, no. 17, pp. 164-174, 2015.

[22] E. Hill, *et al.*, "Identification and removal of laser-induced noise in photoacoustic imaging using singular value decomposition," *Biomedical Optics Express*, vol. 8, no. 1, pp. 68-77, 2017.

[23] X. Zhang, Z. Xu, and N. Jia, "De-noising of 3D magnetic resonance images by using higher-order singular value decomposition," *Medical Image Analysis*, 2015.

[24] S. Yeganli, H. Demirel, and R. Yu, "Noise removal from MR images via iterative regularization based on higher order singular value decomposition," *Signal Image & Video Processing*, pp. 1-8, 2017.

[25] R. Mazhar and P. D. Gader, "EK-SVD: Optimized dictionary design for sparse representations," in *Proc. 19th International Conf. on Pattern Recognition*, 2008, pp. 1-4.

[26] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *SIAM Multiscale Modeling and Simulation*, vol. 7, no. 1, pp. 214-241, 2008.

[27] J. Mairal, F. Bach, and J. Ponce, "Non-local sparse models for image restoration," in *Proc. IEEE International Conf. on Computer Vision*, 2009, pp. 2272-2279.

[28] Z. Cai and J. Lai, "An overcomplete learned dictionary based image de-noising method," *ACTA Electronica Sinica*, vol. 37, no. 2, pp. 347-350, 2009.

[29] Q. Xiao, X. Ding, and S. Wang, "Image denoising based on adaptive over-complete sparse representation," *Chinese Journal of Scientific Instrument*, vol. 30, no. 9, pp. 1886-1890, 2009.

[30] Y. Nie, Y. Zhang, and G. Tang, "Random noise suppression algorithm for seismic data based on dictionary learning," *Mathematics in Practice and Theory*, vol. 47, no. 9, pp. 123-128, 2017.

[31] Y. M. He, T. Gan, and W. F. Chen, "Multi-stage image de-noising based on correlation coefficient matching and sparse dictionary pruning," *Signal Processing*, no. 92, pp. 139-149, 2012.

[32] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process*, no. 17, pp. 53- 69, 2008.

[33] M. Protter and M. Elad, "Image sequence de-noising via sparse and redundant representations," *IEEE Trans. Image Process*, vol. 18, no. 1, pp. 27-35, 2009.

[34] L. R. Neira, "Dictionary redundancy elimination," in *Proc. IEEE Proceedings-Vision, Image and Signal Processing*, vol. 151, no. 1, pp. 31-34, 2004.

[35] Q. Guo, Z. Zeng, and C. Yu, "Seismic random noise suppression based on the high-precision dictionary learning algorithm," *Geophysical and Geochemical Exploration*, vol. 41, no. 5, pp. 907-913, 2017.

[36] G. Tang, J. Ma, and H. Yang, "Seismic data denoising based on learning-type overcomplete dictionaries," *Applied Geophysics*, vol. 9, no. 1, pp. 27-32, 2012.

[37] G. Hu, *Modern Signal Processing Tutorial*, Beijing: Tsinghua University Press, 2004.

[38] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform Theory*, vol. 38, no. 2, pp. 617-643, 1992.

[39] Y. Xu, "Wavelet transform domain filters a spatially selective noise filtration technique," *IEEE Trans. Image Processing*, vol. 3, no. 6, pp. 747-758, 1994.

[40] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613-627, 1995.

[41] Y. Wang, X. Cao, and J. Zhang, "Detection and analysis of all-day atmospheric water vapor Raman lidar by using wavelet de-noising algorithm," *Acta Optica Sinica*, vol. 38, no. 2, p. 0201001, 2018.

[42] X. Xu, G. Qu, and Y. Zhang, "Ground roll separation of seismic data based on morphological component analysis in two dimensional domain," *Applied Geophysics*, vol. 13, no. 1, pp. 116-126, 2016.

[43] Y. Chen, P. Zhang, and B. Li, "De-noising of mechanical vibration signals using quantum-inspired adaptive wavelet shrinkage," *Shock and Vibration*, pp. 1-7, 2014.

[44] W. Li, Q. Ma, and Z. Song, "Application of wavelet transform to the analysis of geoelectric field data," *Acta Seismologica Sinica*, vol. 35, no. 1, pp. 26-35, 2013.

[45] W. Liu, H. Chen, and X. Cai, "Signal de-noising algorithm based on dual-tree complex wavelet transform," *Fire Control & Command Control*, no. 12, pp. 84-87, 2014.

[46] N. Luo, L. Wang, and J. Wang, "De-noising of seismic signals based on wavelet transform modulus maximum," *Earthquake Research in Shanxi*, vol. 162, no. 2, pp. 12-15, 2015.

[47] C. Zhang, "Realization and improvement of the modulus maximum de-noising method based on wavelet transformation," *Journal of Nanjing University of Posts and Telecommunications (Natural Science)*, vol. 29, no. 1, pp. 74-77, 2009.

[48] H. Xu, S. Sun, and Z. Gui, "Detection of sub-seismic fault footprint from signal-to-noise ratio based on wavelet modulus maximum in the tight reservoir," *Journal of Applied Geophysics*, vol. 114, no. 3, pp. 259-262, 2015.

[49] S. Li, Y. Zhou, and Y. Zhou, "An adaptive wavelet shrinkage de-noising algorithm for low altitude flying acoustic targets," *Journal of Vibration and Shock*, vol. 36, no. 9, pp. 153-158, 2017.

[50] R. Shao, W. Hu, and J. Li, "Multi-faultfeature extraction and diagnosis of gear transmission system using time-frequency analysis and wavelet threshold de-noising based on EMD," *Shock and Vibration*, no. 20, pp. 763-780, 2013.

[51] A. Peyman, "Improving hydrological process modeling using optimized threshold-based wavelet de-noising technique," *Water Resources Management*, no. 30, pp. 1701-1721, 2016.

[52] Q. Miao, L. Shan, and Y. Qiu, "The application of entropy in the segmentation of data set," *Computer Knowledge and Technology*, no. 3, pp. 1193-1194, 2007.

[53] Q. Xia and R. Wang, "Mass data partition for rough set on attribute reduction algorithm," *Computer Technology and Development*, vol. 20, no. 4, pp. 5-11, 2010.

[54] J. Wang and G. Zhang, "A new segmentation method of large scale image datasets," no. 4, pp. 61-68, 2015.

[55] S. Li, *Wavelet Transform and Its Application*, Beijing: Higher Education Press, 1997.

Chuanchao Zhang was born in China in 1976. In 2000, he had graduated from Northwest Polytechnic University, obtained bachelor degree in aircraft design. In 2008, he had graduated from Civil Aviation University of China, obtained master degree in human-machine and environmental engineering. At present, he is studying in Wuhan University of Technology for doctoral degree in information and communication engineering. Main research directions: aircraft design, aircraft health management and big data mining.

He is a senior engineer, and has participated many aircrafts design work, and has participated project researches for many aircrafts informatization, standardization, integrated logistics support, and has participated in a number of topics researches related civil aviation. He has been an aircraft design engineer, assistant editor-in-chief of International Aviation Journal, executive editor of Aerospace Science and Technology Journal, *editor-in-c*hief of UV Journal, etc. He has published several papers about aircraft monitoring and data mining in domestic and foreign journals, such as, such as "Extracting algorithm for diagnostic parameters of aero-engine based on fuzzy-rough sets" (Beijing: Journal of Aerospace Power, 2008). He has published many books, such as Comprehensive Unmanned System of the World (Beijing: Aviation Industry Publication, 2015).

Mr. Zhang is a member of the Chinese Aeronautical Society and a member of the Research Association of the China Academy of Low-altitude Industry Economics.