Processing method of EMD endpoint effect based on SVRM extension

Kai Chai[1], Shao-Wei Feng[2]
College of Naval Architecture and Ocean, Naval University of Engineering, Wuhan, 430033, China
[1]Corresponding author
E-mail: [1]*chaikai0805@163.com*, [2]*1787288567@qq.com*

Check for updates

Abstract. Aiming at the problem of endpoint effect in empirical mode decomposition (EMD), the application method of support vector regression machine (SVRM) in EMD extension data prediction is studied. Firstly, the basic principle, data extension method and parameter setting of SVRM are introduced. Secondly, several application methods of SVRM in EMD extension are studied to analyze and verify the operational efficiency and decomposition accuracy characteristics of each method respectively. Finally, the proposed extension method based on SVRM extreme value point prediction can greatly improve the operation efficiency of SVRM long time extension. The simulation signal analysis shows that the SVRM extreme point prediction extension method not only improves the accuracy and reliability of EMD decomposition, but also effectively inhibits the end-point effect phenomenon, significantly reduces the SVRM extension time, and improves the practicability of EMD method.

Keywords: empirical modal decomposition, support vector regression machine, endpoint effect, data extension.

1. Introduction

Empirical mode decomposition (EMD) in the processing of nonlinear, non-stationary signals has unique adaptability, so that it has been in-depth research and application, but the endpoint effect has been restricting the spread of EMD application in seismic signals, structural analysis and mechanical fault diagnosis field, but the endpoint effect has been restricting the spread of EMD application [1].

The extension method based on the original data is based on the data itself, which is fast and suitable for long time extension. Huang et al. found the problem of endpoint effect when the EMD method was proposed, proposed the characteristic wave method and applied for a patent [2]. Mirror closed extension method is simple in operation and efficient in the extension of long data. However, the truncated signal at the extreme value will cause partial information loss, so it is not suitable for the processing of short data, and it is not effective when the signal symmetry is poor [3]. The sinusoidal matching method alleviates the problem of end-point effect to a certain extent, but in practical application, the constructed waveform is difficult to reflect the change trend of the signal, and its practicability is poor [4]. The waveform matching method maintains the change trend of the original signal well, and has a good effect on the signal with regular waveform and strong periodicity [5]. However, for the signal with bad periodicity, it may be difficult to find the wavelet with high matching degree. The algorithm based on data prediction has high general accuracy and good extension smoothness, but the application range of EMD is affected due to the large amount of computation [6]. The prediction algorithm based on neural network has a high accuracy, but it requires a lot of training samples and takes too long learning time to carry out real-time online processing [7]. The adaptive autoregressive model processing method has high operational efficiency, but the autoregressive model itself belongs to linear operation, and its application effect to non-stationary signals is poor [8]. On the basis of summarizing the principle of endpoint effect in EMD decomposition, this paper studies the application of Support vector

machine regression (SVRM) in EMD extension data prediction based on the problem of endpoint effect in empirical mode decomposition (EMD).

2. Overview of SVRM theory

2.1. Data series extension method of SVRM

After constructing the regression training model of SVRM, the data points can be extended forward and backward by using it. The following is an example to introduce the extension method and steps of SVRM:

(a) Construct the training set according to the existing data sequence.

For a given data sequence $s(1), s(2),…, s(N)$, where N is the number of sampling points in the data column. Firstly, the number of training samples l is determined, and a training set $L = \{(x_1 ,y_1), … (x_l, y_l)\}$, where $x_i = [s(i) \quad s(i + 1) \cdots \cdots \quad s(N - l + i - 1)]^T$, $y_i = s(N - l + i)$, $1 \leq i \leq l$.

(b) Construct regression model.

The regression model is constructed according to the above method.

(c) Predict the first sequence value.

The first predicted value outside the boundary can be obtained:

$$s(N + 1) = \sum_{i=1}^{l} \left(\overline{a}_i^* - \overline{a}_i\right) k(x_i, x_{l+1}) + \overline{b}, \quad x_{l+1} = [s(l + 1) \quad s(l + 2) \cdots s(N)]^T \tag{1}$$

(d) Step by step iteration to obtain predictive sequence values.

Then let $s(N + 1)$ as the new boundary point of the original data, the extension value of the second data sequence $s(N + 2)$ can be obtained, and so on. According to the required extension length M, the whole extension sequence can be obtained $s(N), s(N + 1),…, s(N + M)$, the value of any extension $s(N + m)$ is:

$$s(N + m) = \sum_{i=1}^{l} \left(\overline{a}_i^* - \overline{a}_i\right) k(x_i, x_{l+m}) + \overline{b}, \tag{2}$$

where, the value of any extension $s(N + m)$ is $x_{l+m} = [s(l + m) \quad s(l + m + 1) \cdots s(N + m - 1)]^T$.

2.2. Setting of extension parameters

The extension precision of SVRM increases with the increase of the number of samples, while the operation efficiency decreases with the increase of the number of samples. Therefore, in order to ensure the extension accuracy, the number of samples N_e must be increased or decreased with the extension length N_s. Set $k = N_e / N_s$. In general, short time extension requires a higher extension accuracy, and k is set as a smaller value. In the case of long time extension, the extension accuracy requirements can be appropriately reduced, and a larger K can be selected to improve the operational efficiency. Specifically, the length between the signal extreme points is taken as the reference, so that:

$$l = \max[\text{indmin}(2) - \text{indmin}(1), \text{indmax}(2) - \text{indmax}(1)]/2,$$
$$N_e = k_e * l, \quad N_s = k_s * l, \tag{3}$$

where, L is the distance between extremum points; $\text{indmin}(i)$ and $\text{indmax}(i)$ are respectively the moment of the ith minimum and maximum value of the signal, k_s is the sample length coefficient, $k_s > 0$.

3. EMD extension algorithm based on SVRM

In order to prevent the endpoint effect of EMD from spreading into the signal as the decomposition progresses, we always hope to obtain the data extension of sufficient length accurately in a relatively short time. In order to ensure the prediction accuracy of data, sufficient sample quantity is required, while a long extension requires a large sample quantity, which will result in low computational efficiency. Therefore, improving the efficiency of SVRM extension in EMD application has become an urgent problem to be solved.

3.1. EMD primary extension algorithm based on SVRM

The primary extension method refers to the IMF which only conducts primary extension of data before EMD decomposition and then truncates the extension part of the obtained decomposition results to obtain the original data. The characteristic of this method is that the whole decomposition process only needs primary extension, the algorithm is simple to implement, and there is no accumulation of error caused by repeated extension. The simulation signal $y_2(t)$ is taken as an example, and the waveform is shown in Fig. 1. The SVRM method is adopted to carry out a long extension, and then EMD decomposition. Sample length N_s and extension length N_e were controlled by adjusting the values of k_s and k_e. The decomposition of can is shown in Fig. 2:

$$y_2(t) = 3\sin(0.08\pi t) + \sin(0.32\pi t) + \cos(0.01\pi t) + 2\sin(0.04\pi t),$$
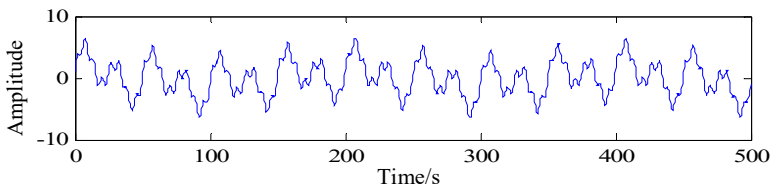$$N = 2000, \quad t = 500s, \quad f_s = 4. \tag{4}$$

Fig. 1. Waveform of simulated signal $y_2(t)$

As shown in Fig. 2(a), when $k_s = 1$ and $k_e = 5$, the sample length $N_s = 16$ and the extension length $N_e = 90$, the decomposition results show large deformation at both ends, and the endpoint effect is not effectively contained. When $k_s = 5$ and $k_e = 5$, the sample length N_s increases to 90. At this time, the decomposition results also show a large deformation at both ends, that is, the endpoint effect is effectively contained without the increase of the sample length. It can be inferred that when $k_e = 5$, the extension length is too short, so the endpoint effect is not effectively contained outside the real signal. When $k_s = 1$ and $k_e = 20$, the extension length N_e increases to 320 points, and the decomposition results are greatly improved. There is no obvious distortion at the end points, but residual term r_5 is generated. It can be inferred that when $k_s = 1$, the regression model cannot accurately predict the signal trend due to the small sample data. When $k_s = 5$ and $k_e = 20$, the sample length increased to 80 points. At this time, a more accurate decomposition result was obtained, but the running time of the program increased to 71.8776 s.

3.2. EMD decomposition and extension simultaneously based on SVRM

In order to avoid the problem of operation efficiency caused by SVRM long time extension, an effective method is to make full use of the accuracy of SVRM short time extension, and adopt appropriate algorithm to avoid long time extension. The effective method is to decompose while extension. The idea is to conduct a short time extension with a length of more than two extreme points before each screening of EMD, and then cut off the extension part after generating the envelope. Since the algorithm needs to be extended before each screening, it only needs to extend

a pair (a maximum value and a minimum value) above the extreme point at both ends of the signal each time. Repeated simulation experiments verify that, under the same conditions, when the two ends extend more than four extreme points respectively, the decomposition results will no longer change significantly. In order to ensure the decomposition accuracy, the method of extending four extreme points at both ends is adopted in this paper. Set $k_s = 4$, then the EMD processing results of $y\_2(t)$ are shown in Fig. 3.
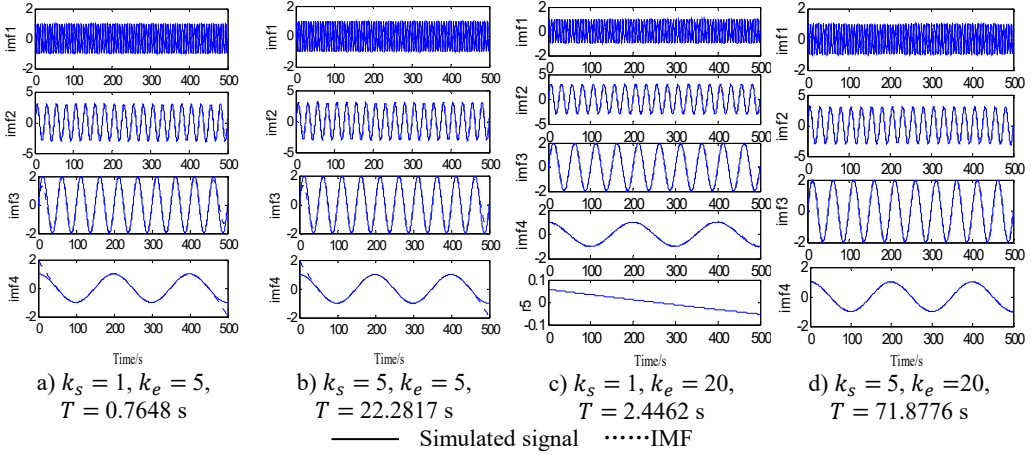

a) $k_s = 1$, $k_e = 5$, $T = 0.7648$ s

b) $k_s = 5$, $k_e = 5$, $T = 22.2817$ s

c) $k_s = 1$, $k_e = 20$, $T = 2.4462$ s

d) $k_s = 5$, $k_e = 20$, $T = 71.8776$ s

—— Simulated signal ······IMF

Fig. 2. Decomposition of simulated signal after one extension

In the Fig. 3, when the sample is 10, residual terms are generated in the decomposition result; when the sample is increased to 20, the decomposition result reaches a very high precision, and at this time, the whole decomposition process takes only 0.9045 s.

It can be seen from Fig. 3 that when $k_s = 1$, due to the small sample number and poor extension accuracy, the endpoint effect is restrained to some extent, but residual terms appear. When $k_s = 2$, a relatively accurate decomposition result was obtained, and the operation time reached 388.2863 s.
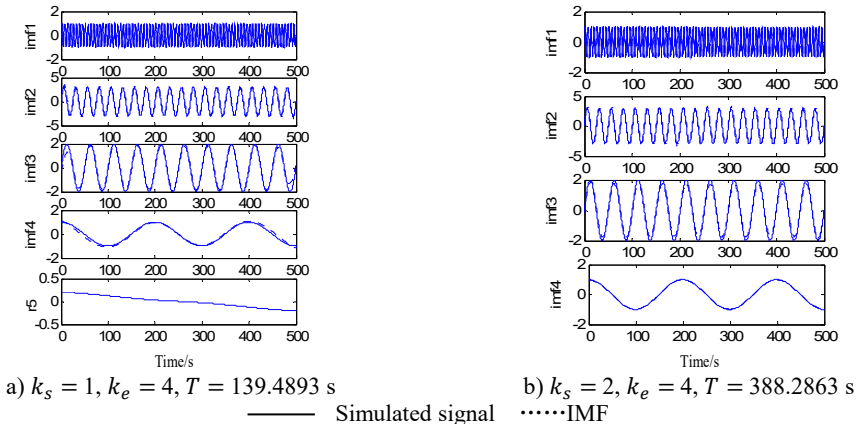

a) $k_s = 1$, $k_e = 4$, $T = 139.4893$ s

b) $k_s = 2$, $k_e = 4$, $T = 388.2863$ s

—— Simulated signal ······IMF

Fig. 3. Decomposition of simulated signal after decomposition and extension simultaneously

3.3. EMD continuation based on SVRM extremum prediction

The extremum forecast and continuation process, SVRM extremum predict continuation of sample sequence is extremum sequence of original signal, so the methods require original signal

have enough number of extreme value (experiments show general may not be less than 6), when the signal extremum number is small, SVRM based on hard less extreme value point accurately predict epitaxial signal extremum position and value. It should be noted that, with the progress of EMD decomposition, the frequency components of the obtained components gradually decrease. If this method is used for side decomposition and side continuation, the phenomenon of inaccurate continuation may occur in the low-frequency stage due to the small number of extreme points. Therefore, this method is more suitable for the previous continuation of decomposition. The decomposition results are shown in Fig. 4.
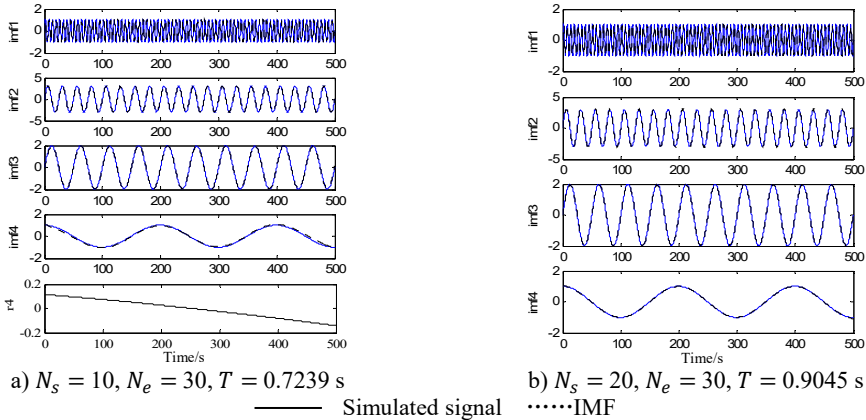

a) $N_s = 10$, $N_e = 30$, $T = 0.7239$ s b) $N_s = 20$, $N_e = 30$, $T = 0.9045$ s

———— Simulated signal ······IMF

Fig. 4. EMD results of continuation prediction based on SVRM extreme value

In order to verify the validity of the effect of the endpoint effect and the practical practicability of the actual hydraulic system, the sampling frequency of the measured hydraulic system is analyzed, and the sampling frequency is 5 kHz, and the sampling point is 2048. The results of EMD continuation based on SVRM extremum prediction results are as shown in the Fig. 5. The weak at both ends, and the similar components of the frequency are not visible in the endpoints, and the model of the IMF component is obviously suppressed.
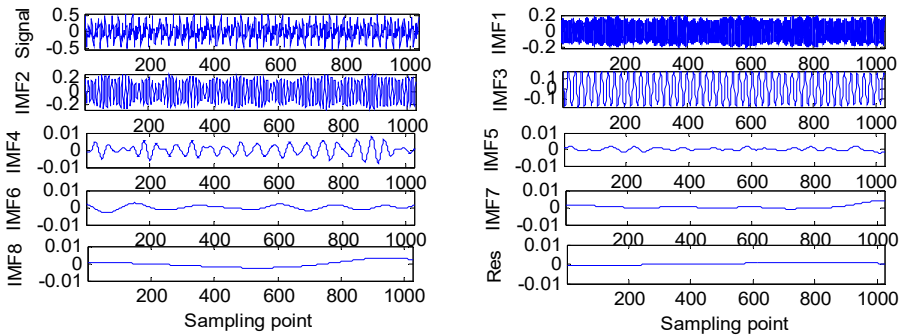

Fig. 5. EMD results of hydraulic fault signal

4. Conclusions

Based on the application of SVRM in EMD continuation, this paper introduces the prediction principle of SVRM, and proposes a method to set the extension length and sample number with the signal extremum scale. The advantages and disadvantages of various SVRM application methods in EMD continuation are analyzed from the Angle of decomposition accuracy and efficiency.

References

[1] **Huang N. E., Shen Zheng, Long S. R., et al.** The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, Vol. 454, Issue 1971, 1998, p. 903-995.

[2] **Huang N. E., Wu M. L., Qu W. D.** Applications of Hilbert-Huang transform to non-stationary financial time series analysis. Applied Stochastic Models in Business and industry, Vol. 19, Issue 3, 2003, p. 245-268.

[3] **Zhao Jinping** Improvement of the mirror extending in empirical mode decomposition method and the technology for eliminating frequency mixing. High Technology Letters, Vol. 8, Issue 3, 2002, p. 40-47.

[4] **Lin D. C., Guo Z. L., An F. P., et al.** Elimination of end effects in empirical mode decomposition by mirror image coupled with sup ort vector regression. Mechanical Systems and Signal Processing, Vol. 31, 2012, p. 13-28.

[5] **Tang B. P., Dong S. J., Song T.** Method for eliminating mode mixing of empirical mode decomposition based on the revised blind source separation. Signal Processing, Vol. 92, Issue 1, 2012, p. 248-258.

[6] **Guhathakurta K., Mukherjee I., Chowdhury A. R.** Empirical mode decomposition analysis of two different financial time series and their comparison. Chaos, Solitons and Fractals, Vol. 37, 2008, p. 1214-1227.

[7] **Yu L., Wang S., Lai K. K.** Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics, Vol. 30, 2008, p. 2623-2635.

[8] **Wu F. J., Qu L. S.** An improved method for restraining the end effect in empirical mode decomposition and its applications to the fault diagnosis of large rotating machinery. Journal of Sound and Vibration, Vol. 314, 2008, p. 586-602.