People's Demecratic Republic of Algeria
Ministry of Higher Education and Scientific research
M'hamed Bougara University, Boumerdes
Institute of Electrical and Electronic Engineering,
Laboratory of Signals and Systems (LSS)

ALGERIAN JOURNAL OF SIGNALS AND SYSTEMS

ISSN : 2543-3792

Title: **Fast Ensemble Empirical Mode Decomposition Using the Savitzky Golay Filter**

Authors: **Wahiba Mohguen and Raïs El'hadi Bekka**
Affiliation: **Department of Electronics, LIS LaboratoryFaculty of Technology, Université de Sétif 1, 19000 Sétif, Algeria.**

Page range: **79- 86**

IMPORTANT NOTICE

This article is a publication of the Algerian journal of Signals and Systems and is protected by the copyright agreement signed by the authors prior to its publication. This copy is sent to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration. Other uses, namely reproduction and distribution, selling copies, or posting to personal, institutional or third party websites are not allowed.

Volume : 1 Issue : 1 (June 2016)

Laboratory of Signals and Systems
Address : IGEE (Ex-INELEC), Boumerdes University, Avenue de l'indépendance, 35000, Boumerdes, Algeria
Phone/Fax : 024 79 57 66
Email : lss@univ-boumerdes.dz ; ajsyssig@gmail.com

©LSS/2016

Fast Ensemble Empirical Mode Decomposition Using the Savitzky Golay filter

WAHIBA MOHGUEN AND RAÏS EL'HADI BEKKA*

Department of Electronics, LIS Laboratory
Faculty of Technology, Université de Sétif 1
19000 Sétif, Algeria
*bekka_re@yahoo.fr

Abstract: Empirical mode decomposition (EMD) is a powerful algorithm proposed to analysis of nonlinear and non-stationary signals. The phenomenon of mode mixing is one of the major disadvantages of the EMD. The Ensemble EMD (EEMD) was introduced to eliminate the mode-mixing effect. The principle of EEMD is to add additional white noise into the signal with many trials. The noise in each trial is different; and the added noise can be completely cancelled out on average, if the number of trials is very high. The number of trials is a high computational load. The improvement on computational efficiency of EEMD is therefore required. In this paper, an improvement on the computing time of the EEMD was proposed by replacing white noise with white noise filtered using Savitzky-Golay (SG) filter. Numerical simulations were performed to demonstrate that such replacement has effectively reduced the number of trials to obtain a noise-free reconstructed signal.

Keywords: EMD, EEMD, Mode mixing, Savitzky-Golay filter.

1. INTRODUCTION

The Empirical Mode Decomposition (EMD) method was introduced by Huang et al. [1] for the analysis of non stationary and/or nonlinear signals. The EMD was designed to adaptively decompose a signal into a set of intrinsic mode functions (IMFs) using a sifting process. One of the major drawbacks of the original EMD is the phenomenon of mode mixing which is defined by the coexistence of oscillations of different time scales in a single IMF or existence of the same oscillation with the same time scale in different IMFs [2]. Mode mixing is often produced by signal intermittent. To overcome the mode mixing problem, a noise assisted data analysis method called the ensemble empirical mode decomposition (EEMD) was recently developed by Wu and Huang [2]. The principle of EEMD is to add white noise into the signal with an ensemble of trials and the mixed signals are repeatedly decomposed by EMD. EEMD defines the true IMF components as the mean of an ensemble of trials, each consisting of the signal plus a white noise of finite amplitude. Indeed, EEMD achieves a significant improvement over the original EMD method for eliminating the mode mixing problem. But theoretically, the added noise can be completely cancelled out on average, if the number of trials would be infinite. Therefore, the very high number of ensemble trails to eliminate the residue noise in the signal reconstruction is synonymous with the very high computational cost. In order to improve the efficiency of the EEMD algorithm, many approaches were proposed [3]-[9].

Yeh et al. [3] expanded the EEMD method into complementary EEMD (CEEMD) by using both positive and negative white noise to the original data (i.e. one positive and one negative) to generate two sets of ensemble IMFs. CEEMD has effectively eliminated residue noise in the IMFs. However, a larger number of ensemble trials are also necessary to yield the root-mean- square (RMS) noise of CEEMD comparable to that of EEMD. Recently, Zheng et al. [4] modified the CEEMD method and developed a faster noise-assisted method called partly ensemble empirical mode decomposition (PEEMD) by performing the original EMD method on the noise signal obtained by an ensemble way and detected by permutation entropy. Although an important

improvement in computation time was achieved by PEEMD, the determination of the permutation entropy threshold is not adaptive and tough.
Zhang et al. [5] proposed a modified EEMD (MEEMD) method to reduce the computational cost of the original EEMD by replacing white noise with a band-limited noise to be added to the signal to be decomposed. But, this improvement is not significant. Recently, a sensitive improvement over the MEEMD was achieved with the right choice of the filter type and its characteristics [6].
A complete ensemble empirical mode decomposition with Adaptive Noise (CEEMDAN) was developed by Torres et al. [7]. In this method, a particular noise was added at each stage of the decomposition and a unique residue was computed to obtain each IMF, and achieved a complete decomposition with no reconstruction error. The CEEMDAN method requires less than half the sifting iterations that EEMD does.
Bekka and Berrouche [8] showed that a significant reduction for the computational complexity in EEMD can be obtained by over-sampling the original signal to reduce the ensemble trials. For very high-frequency signals, this method can lead to physically unrealizable sampling frequencies.
Very recently, an adaptively fast EEMD method combined with complementary EEMD was proposed by Xue, et al. [9] to resolve the problem of high computational cost.
In this paper an improved EEMD method, namely SGEEMD, was proposed. In this method, a significant improvement of computational efficiency was obtained with a white noise filtered using Savitzky-Golay (SG) filter. Numerical evaluation on a test signal showed the effectiveness of the SGEEMD method. A comparative study with the original EEMD showed that the proposed method allowed reducing the computation time by approximately 90%.

2. EMD AND EEMD ALGORITHMS

EMD Algorithm

EMD is an adaptive method to decompose asignal $x(t)$ into a set of IMFs. The IMFs must satisfy the following two conditions:

(1) The number of maximum must equal the number of zeros or differ at most by one.
(2) In each period, it is necessary that the signal average is zero.

The EMD algorithm (the sifting process of extracting IMFs) consists of the following steps [1]:

1. Identify local maxima and minima in $x(t)$ to construct the upper and lower envelopes respectively using cubic spline interpolation.
2. Calculate the mean envelope, $m(t)$, by averaging the upper and lower envelopes.
3. Calculate the temporary local oscillation $h(t) = x(t) - m(t)$.
4. Calculate the average of $h(t)$, if average is close to zero, then $h(t)$ is considered as the first IMF, named $IMF_i(t)$, otherwise, repeat steps (1)–(3) while using $h(t)$ for $x(t)$.
5. Calculate the residue $r(t) = x(t) - c_i(t)$.
6. Repeat the procedure from steps (1) to (5) using $r(t)$ for $x(t)$ to obtain the next IMF and residue.

The decomposition process stops when the residue $r(t)$ becomes a monotonic function or a constant that no longer satisfies the conditions of an IMF.

$$x(t) = \sum_{i=1}^{N} IMF_i(t) + r_N(t) \tag{1}$$

The sifting process is continued until the last residual is either a monotonic function or a constant.

EEMD Algorithm

The true IMFs in the EEMD are defined as the mean of an ensemble of trials. Each trial consists of the signal and the added white noise of finite amplitude. This noisy signal is then decomposed by EMD into noisy IMFs. In fact, the true IMFs are obtained by averaging the IMFs of the same order. Since the noise in each trial is different from other trials, noise added to the true IMFs is canceled out in the ensemble mean of relatively large number of trials [3]. For a number of ensemble trials given N_t, the EEMD algorithm consists of the following steps [2]:

1. Add a white noise $n(t)$ to the target signal $x(t)$: $x_j(t) = x(t) + n_j(t),\ 1 \le j \le N_t$

2. Decompose the noisy signal $x_j(t)$ by EMD algorithm into IMFs.

3. Repeat step 1 and step 2 until the end of number of ensemble trials, but with different added white noise each time.

4. Estimate the true IMFs of EEMD noted IMF_{iEEMD} by averaging the IMFs of the same order:

$$IMF_{iEEMD}(t) = \frac{1}{N_t}\sum_{j=1}^{N_t} IMF_j(t), \quad 1 \le i \le N \tag{2}$$

where N denotes the IMF number.
In the SGEEMD algorithm, $n(t)$ is replaced par $n_{SG}(t)$: a white noise filtered by Savitzky-Golay (SG) filter.

3. SG FILTER

The data smoothing is to replace each data point by some kind of local average of surrounding data points. The averaging can reduce the level of noise without biasing the value obtained. In this work a particular type of low-pass filter, well-adapted for data smoothing termed Savitzky-Golay filter was used to eliminate the residue noise by assuming that relatively distant data points have some important redundancy that can be used to reduce the level of noise.
Given N samples of a white noise $\eta(n)$ $(\eta(0), \eta(1), \cdots, \eta(N-1))$ and a window size of $2M+1$ points (frame size) centered at the central point of the interval, the smoothed value of $\eta(i)$ by Savitzky-Golay filter is given by:

$$c(i) = \sum_{j=-M}^{M_t} g(j)\eta(i+j) \tag{3}$$

where $g(j)$ are filter coefficients.
$\hat{c}(i)$ is the average of the data points $\eta(i-M), \cdots, \eta(-1), \eta(0), \eta(1), \cdots, \eta(i+M)$. The concept of Savitzky-Golay filtering is to find the coefficients $g(j)$ that preserve higher moments within the moving window by a higher order polynomial [9]. For a polynomial order and for each point $\eta(i)$, the coefficients $g(j)$ must be determined optimally to minimize the least-squares error in fitting a polynomial to all $2M+1$points in the moving window such that the $\hat{c}(i)$ *is the* corresponding value of the polynomial curve at position *i*.

4. RESULTS AND DISCUSSION

In order to evaluate the accuracy of the SGEEMD, a simulated signal $x(t)$ consisting of widely disparate scales, i.e. permanent component of low frequency $(100\ Hz)$ and of a transient component of high frequency $(1500\ Hz)$ was used to show the resolution of the mode mixing problem and the reduction of residue noise in the reconstructed signal. The signal is described as (Fig.1):

$$x(t) = \sin(2\pi f_1 t) + \lambda \sin(2\pi f_2 t)\left(e^{-\frac{(t-t_0)^2}{\sigma}} + e^{\frac{-(t-t_1)^2}{\sigma}} + e^{\frac{-(t-t_2)^2}{\sigma}} + e^{\frac{-(t-t_3)^2}{\sigma}}\right), 0 \le t \le 0.055\ s \tag{4}$$

With $t_0 = 0.0125\ s$, $t_1 = 0.0225\ s$, $t_2 = 0.0325\ s$, $t_3 = 0.0425\ s$, $\sigma = 10^{-6}$, $\lambda = 0.1$, $f_1 = 100\ Hz$ and $f_2 = 1500 Hz$.
The test signal was sampled at $f_e = 10000 Hz$.

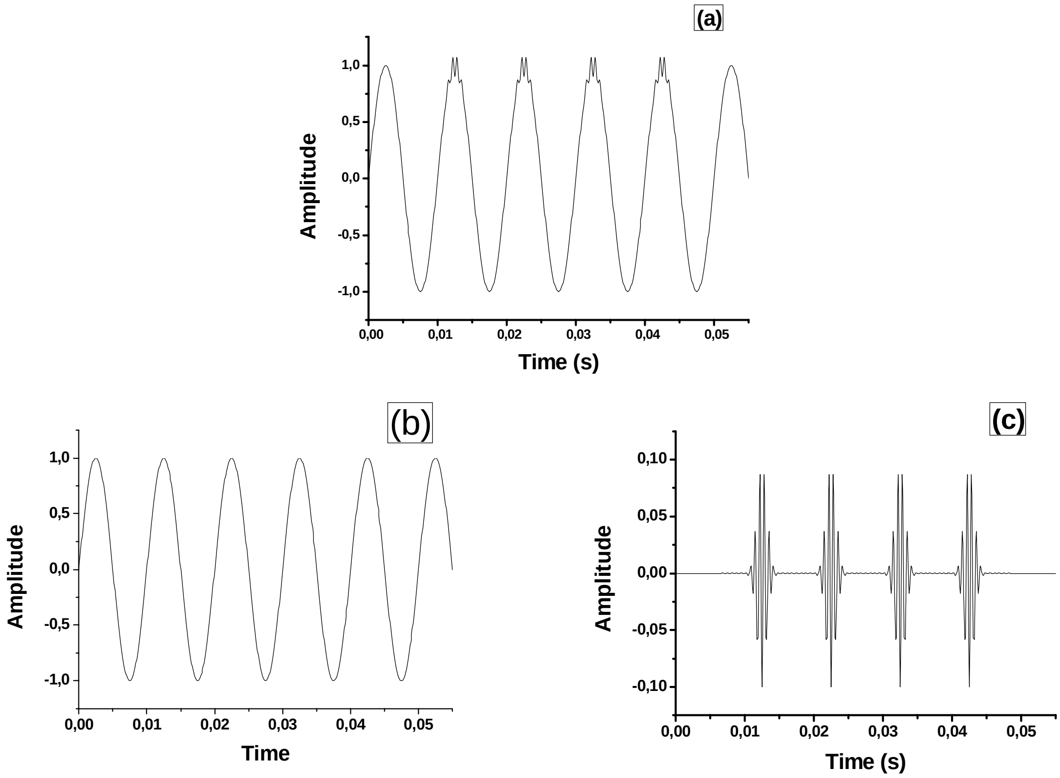


Fig. 1 (a) The simulated signal $x(t)$ and its components: (b) The components of low-frequency and (c) The intermittent of high frequency.

EMD decomposes the signal $x(t)$ into three IMFs (Fig. 2). The IMF1 in "Fig. 2" shows mode-mixing deficiency where both high frequency and low frequency components are mixed. To solve the problem of mode mixing, the EEMD was proposed [2], where Gaussian white noises with finite amplitude are added to the original signal during the decomposition process. However, to overcome the problem of mode mixing, the noise amplitude must be controlled. A suitable choice of the amplitude of the noise is obtained by using the signal to noise ratio *SNR* [5]. Figure 3a shows the relationship between *SNR* and IMFs resulting from the EEMD decomposition of signal $x(t)$. Positive values represent the number of redundant IMFs, the value of 0 indicates that there is no redundant IMF, and the value of -1 corresponds to mode mixing. Therefore, "Fig. 3a" shows that *SNR* within the range of 37-45 *dB* provides a good EEMD decomposition without mode mixing and also without redundant IMF. The number of ensemble trials must be theoretically infinite to fully remove the impact of the white noise added to the signal. Thus, the number of ensemble trials is also a parameter that must be controlled. In this study, the number of ensemble trials was determined by the correlation coefficient between IMF1 and the corresponding component in the simulated signal. Fig. 3b shows when the number of ensemble trials is 100, the corresponding correlation coefficient is 0.99 which an adequate threshold value to determine the number [5], [8].
The EEMD decomposition result of signal $x(t)$ is shown in "fig. 4", where the $SNR = 37dB$ and $N_t = 100$. As a result, mode mixing is effectively eliminated by the EEMD process. It should be noted however that after 100 trials that IMF1 still contains a very low residue that is visible in "Fig.4a". Although the EEMD method has solved the mode mixing problem, the large number of ensemble trails to reduce the added noise increases the computational time. It has been shown that a solution to the problem of high computational cost of the EEMD method can be obtained by replacing the white noise by a band-limited noise [5], [6]. Such substitution has reduced the number of ensemble trials required to obtain free-noise IMFs, especially by appropriate choice of the filter type and its order [6].
In this work, an improvement of computing time is obtained by filtering the white noise through the SG filter. The parameters of the SG filter were 41 for size, and 3 for the order. The relationship

between the *SNR* and the SGEEMD decomposition results for the test signal is given in "Fig. 5". This figure shows that *SNR* within the range of 30-42 *dB* provides a good SGEEMD decomposition without mode mixing and also without redundant IMF. Figure 6 shows the decomposition results using SGEEMD. The high frequency component has been easily identified in IMF1 after only 10 trials, instead of the 100 trials required previously. In addition, it can be appreciated in Fig. 6a that the residue in the IMF1 was completely eliminated. So, the proposed method provides a sharper decomposition than that of EEMD Method. The performance of proposed new method was also evaluated using Root Mean Squared Error (*RMSE*) and the correlation coefficient between the IMF1 and the corresponding high-frequency component in the simulated signal $x(t)$. Results are also compared with those of the reference [6], where the white noise was filtered using a low-pass Elliptic filter of 4^{th} order. The table 1 clearly shows the advantage of using a white noise filtered by SG filter. It is seen from table 1 that the percentage improvement of the computational efficiency is 66.66 % compared to MEEM method with 4th order elliptic low-pass filter [6] and 90% compared to the original EEMD method.

In summary, the method here proposed provides an exact reconstruction of the original signal by summing the IMFs with a lower computational cost.

Table 1 Comparative Study Between EEED ,MEEMD (4th order ellipticallow-pass filter) and SGEEMD.

	EEMD	**MEEMD (4th order elliptical low-pass filter)**	**SGEEMD**
***SNR* (*dB*)**	37 *dB*	10 *dB*	39 *dB*
Correlation coefficient	0,99	0,99	0.99
N_t	100	30	10
$RMSE_{IMF1}$	0,0015	0,0019	0.001

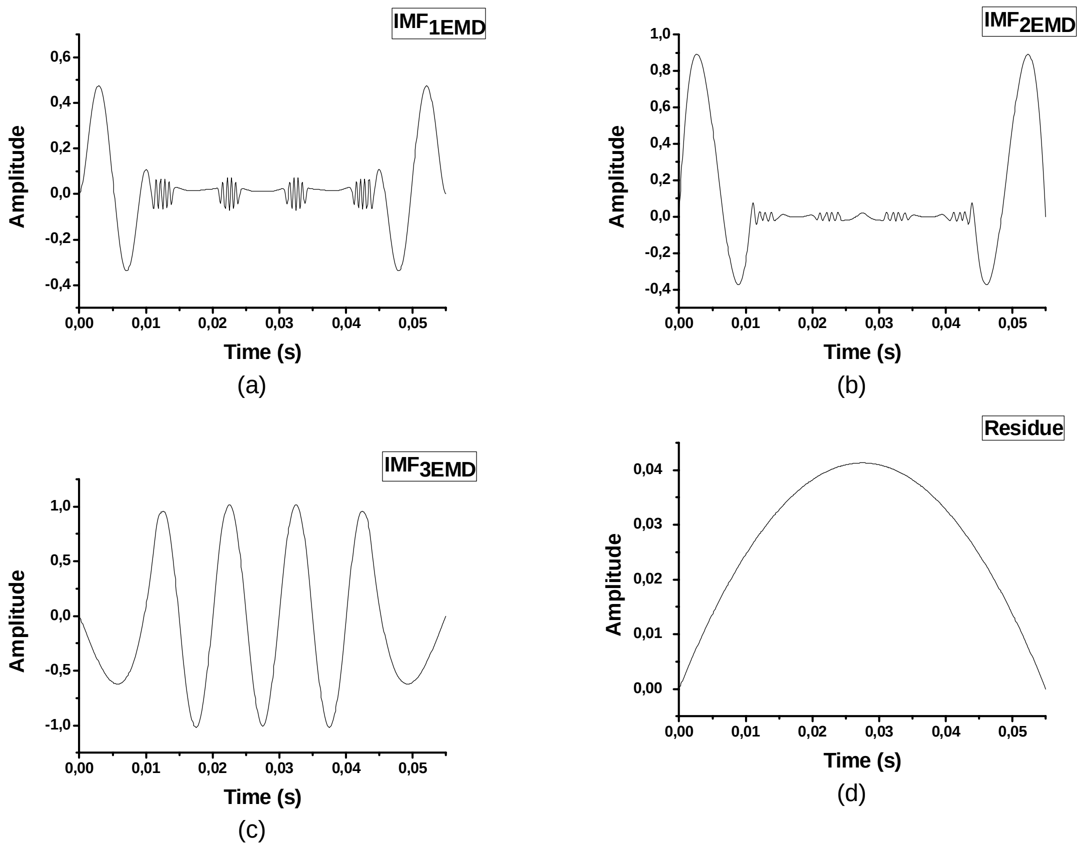
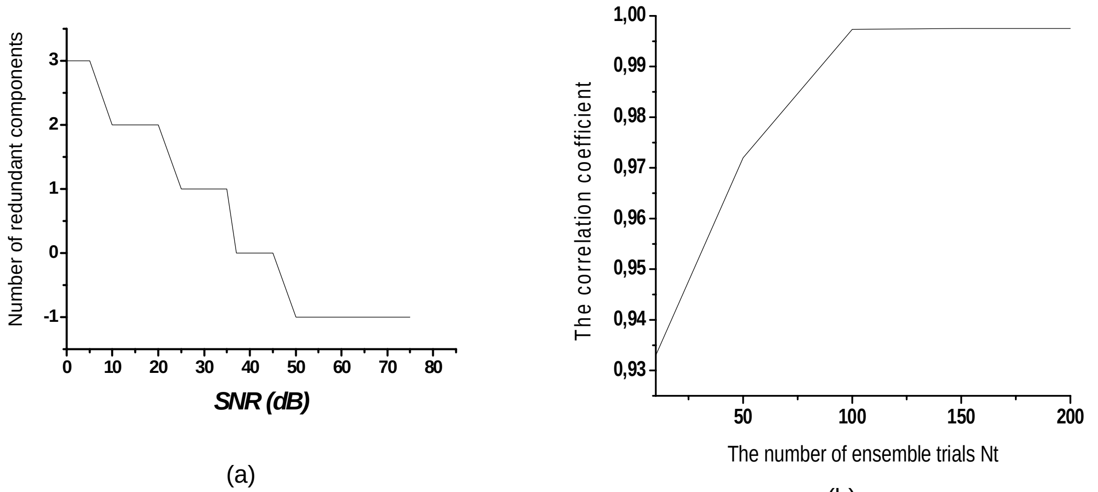


Fig. 2 Decomposition of signal *x*(*t*) by EMD method. Mode mixing is observed in $IMF_{1EMD\ in\ (a)}$.

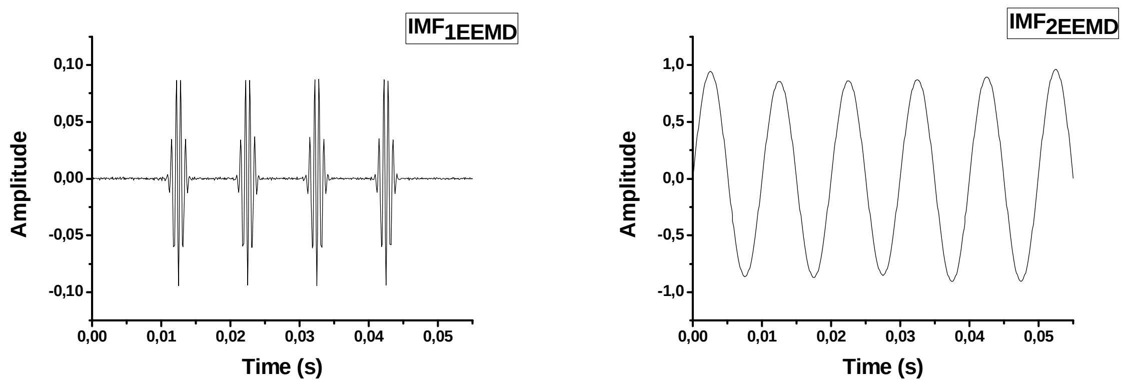


(a)

(b)

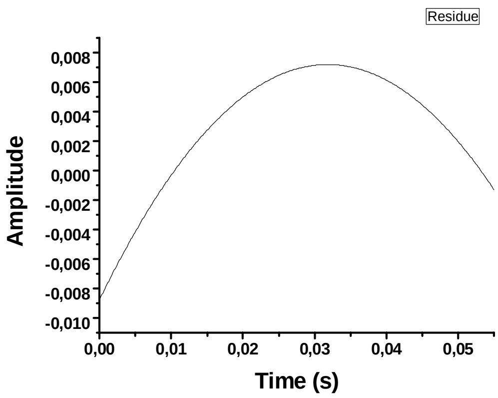
Fig. 3 (a) Relationship between *SNR* and the decomposition results of the signal $x(t)$ by EEMD method $N_t = 100$.

(b) Relationship between the correlation coefficient and the number of ensemble trials in EEMD method.



(a)

(b)



(c)

Fig. 4 Decomposition of signal $x(t)$ by EEMD method with $SNR = 37dB$ *and* $N_t = 100$.

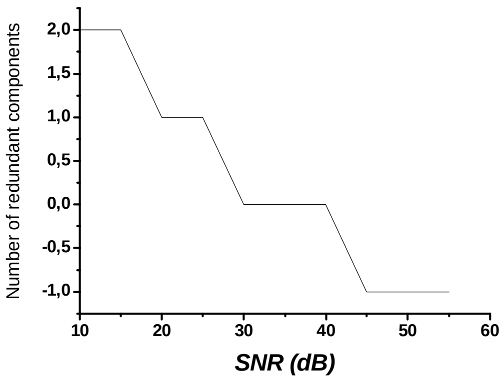
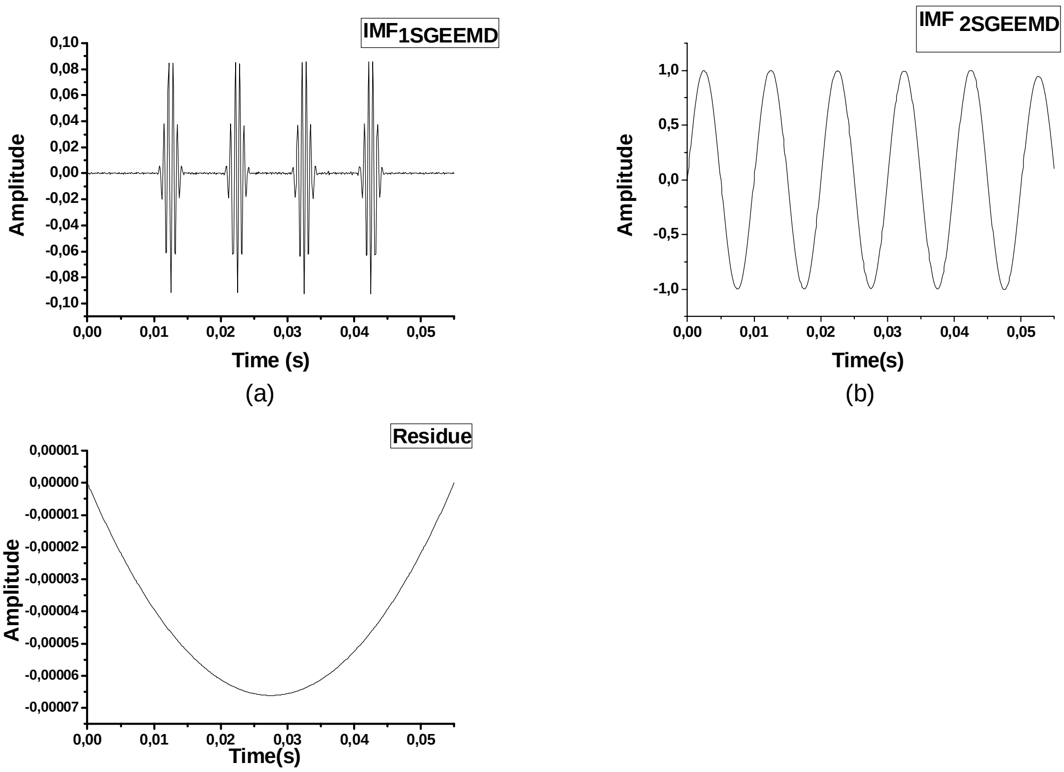


Fig. 5 Relationship between IMFs and *SNR* of the signal $x(t)$ by SGEEMD method.



(a) (b)

(c)

Fig. 6 Decomposition of signal $x(t)$ by SGEEMD with $SNR = 39dB$ *and* $N_t = 10$.

5. CONCLUSION

In this paper an improved EEMD method, called SGEEMD, has been proposed to reduce the residue noise in the IMFs and thus to allow an exact reconstruction of the original signal whilst also reducing the computational cost of the original EEMD method. This new method uses white noise filtered by SG filter instead of white noise to improve processing efficiency of original EEMD by reducing the number of ensemble trials. In SGEEMD, the high frequency component can be easily identified in the first IMF after 10 ensemble trials, instead of the 100 trials needed by the original method. This represents an improvement of the computational efficiency of approximately 90%. In addition the *RMSE* derived from SGEEMD to decompose the test signal was smaller than that obtained with the original EEMD.

Finally, The results showed that the amplitude of the added white noise must be chosen appropriately to ensure EEMD performance.

. References

[1] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen Tung., H. H. Liu, “The empirical mode decomposition and the hilbert spectrum for non-linear and non-stationary time series analysis,” Proceedings of the Royal Society of London, 454, pp. 903–995, 1998.

[2] Z. Wu, N. E. Huang, “Ensemble empirical mode decomposition: a noise-assisted data analysis method,” Adv. Adapt. Data Anal., vol. 1 (1), pp. 1–41, 2009.

[3] J. R. Yeh, J. S.Shieh, “Complementary ensemble empirical mode decomposition a novel noise enhanced data analysis method,” Adv. Adapt. Data Anal., vol. 2 (2), 135–156, 2010.

[4] J. D. Zheng, J. S.Cheng, Y.Yang, “Partly ensemble empirical mode decomposition: an improved noise-assisted method for eliminating mode mixing,” Sign. Process. 96, 362–374, 2014.

[5] J. Zhang, R. Yan, R. X. Gao, Z. Feng Z., “Performance enhancement of ensemble empirical mode decomposition,” Mechanical Systems and Signal Processing, vol. 24 (7), pp. 2104-2123, 2010.

[6] W. Mohguen, R. E. Bekka, "Improvement of ensemble Empirical Mode Decomposition by a band-limited white noise", 2015 4th International Conference on Electrical Engineering (ICEE), Boumerdes, Algeria, pp. 1 - 4, 13-15 Dec. 2015.

[7] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise, ” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4144–4147, 2011.

[8] R. E. Bekka, and Y. Berrouche, “Improvement of ensemble empirical mode decomposition by over-sampling” Adv. Adapt. Data Anal., **5**, (10 pages) 2013.

[9] X. Xue,J. Zhou, Y. Xu,W. Zhu,C. Li, “An adaptively fast ensemble empirical mode decomposition method and its applications to rolling element bearing fault diagnosis”, Mech.Syst.Sig.Process. vol. 62–63, pp. 444–459, 2015.

[10] W. H. Press, B. P. Flannery, S. A Teukolsky, W. T. Vetterling ,“*Numerical Recipes in C: The Art of Scientific Computing*.” New York, NY: Cambridge University Press, 1988.