

Ensemble Empirical Mode Decomposition: An adaptive method for noise reduction

Megha Agarwal¹, R.C.Jain²

¹(ECE, Jaypee Institute of Information and Technology, Noida, India)

²(ECE, Jaypee Institute of Information and Technology, Noida, India)

Abstract: Empirical mode decomposition (EMD), a data analysis technique, is used to denoise non-stationary and non-linear processes. The method does not require any pre & post processing of signal and use of any specified basis functions. But EMD suffers from a problem called mode mixing. So to overcome this problem a new method known as Ensemble Empirical mode decomposition (EEMD) has been introduced. The presented paper gives the detail of EEMD and its application in various fields. EEMD is a time-space analysis method, in which the added white noise is averaged out with sufficient number of trials; and the averaging process results in only the component of the signal (original data). EEMD is a truly noise-assisted data analysis (NADA) method and represents a substantial improvement over the original EMD.

Keywords –Data analysis, Empirical mode decomposition, intrinsic mode function, mode mixing, NADA,

I. INTRODUCTION

Data analysis is an essential part in pure research and practical applications. Basically it is defined as a process of evaluating data using analytical and logical reasoning to examine each component of the data provided. As it is well known fact that linear and stationary processes are easy to analyze, but the real world signals are mostly non-linear and non-stationary in nature. Analysis of such time varying signals is not an easy process. This gives rise to breakdown the process (under consideration) into individual components and analyze each component separately. Breaking out a complex process into separate components called decomposition. There exist a number of time frequency (TF) representation methods of time domain signal such as Fourier Transform (FT), Short Time Fourier Transform (STFT), Wavelet transform, Wigner Ville distribution, and evolutionary spectrum. Out of these FT, STFT, and Wavelet are widely used which are described as follows:

- a) Fourier Transform: Historically, Fourier spectrum analysis has provided a general method for examining the global energy-frequency distribution. Fourier analysis has dominated the data analysis efforts soon after its introduction because of its prowess and simplicity. The Fourier transform belongs to the class of orthogonal transformations that uses fixed harmonic basis functions. The Fourier transform result can be shown as a decomposition of the initial process into harmonic functions with fixed frequencies and amplitudes. But the FT is valid under extremely general conditions, i.e. the system must be linear; and the data must be strictly periodic or stationary; otherwise the resulting spectrum will make little physical sense.
- b) STFT: It is a limited time window-width Fourier spectral analysis. Since it relies on the traditional Fourier spectral analysis, it is assumed that the data has to be piecewise stationary. So in case of non-stationary signals, it has limited usage.
- c) Wavelet transform: To avoid constraints associated with non-stationarity of the initial sequence, a wavelet transform is used. Like the Fourier transform, it performs decomposition in a fixed basis of functions. But unlike FT it expands the signal in terms of wavelet functions which are localized in both time and frequency. But for practical purposes, it would be good to have a transform that would not only allow dealing with non-stationary processes but would also use an adaptive transform basis determined by initial data.

II. EMPIRICAL MODE DECOMPOSITION

Empirical Mode Decomposition has been introduced by Huang [1] for analyzing non-linear and non-stationary signal. EMD effectively overcome the limitations of above described methods. It is an iterative process which decomposes real signals x into elementary signals (modes) [5]. In this method, first the signal is decomposed into a number of IMF. For this the condition of IMF should be verified which are given below:

- a. In the whole data set, the number of extrema and the number of zero crossings must be either equal or differ at most by one.
- b. At any data point, the mean value of the envelope defined using the local maxima and the envelope defined using the local minima are zero.

II.1 Sifting procedure

- a. Compute a mean envelope $m_1(t)$ of the signal $x(t)$.
 - b. Let $h_1(t) = x(t) - m_1(t)$ be the residue.
 - c. If $h_1(t)$ is an IMF, STOP; else, treat $h_1(t)$ (with its extrema) as a new signal to obtain $h_{1,1}(t)$.
 - d. If $h_{1,1}(t)$ is an IMF, STOP; else, continue the same process
 - $h_{1,1}(t) = h_1(t) - m_{1,1}(t)$
 - ...
 - $h_{1,k}(t) = h_{1,k-1}(t) - m_{1,k}(t)$.
- Generally, after a finite number k_1 times, $h_{1,k_1}(t)$ will be an IMF, denoted by $IMF_1(t)$, the first IMF. Set $r_1(t) = x(t) - IMF_1(t)$. And repeat the sifting procedure:
- $r_2(t) = r_1(t) - IMF_2(t)$
 - ...
 - $r_n(t) = r_{n-1}(t) - IMF_n(t)$

The process ends when r_n has at most one extrema, where n is the total number of decomposed IMF. Thus $x(t)$ is decomposed into finitely many IMFs.

$$x(t) = \sum_{i=1}^n IMF_i(t) + r_n(t)$$

II.2 When does the sifting stop?

To guarantee that the IMF components retain enough information of both amplitude and frequency modulation, a criterion is used to stop the sifting process. This can be accomplished by limiting the size of the standard deviation, SD, computed from two consecutive sifting results as:

$$SD(i) = \frac{\sum_t |IMF_{j,i-1}(t) - IMF_{j,i}(t)|^2}{\sum_t IMF_{j,i-1}^2(t)}$$

III. ENSEMBLE MODE DECOMPOSITION

When a signal contains intermittency the EMD algorithm described above may encounter the problem of mode mixing. Frequent appearance of mode mixing, is defined as a single Intrinsic Mode Function (IMF) either consisting of signals of widely disparate scales, or a signal of a similar scale residing in different IMF components. The intermittence could not only cause serious aliasing in the time-frequency distribution, and also make the individual IMF devoid of physical meaning [3].

To overcome this limitation a new noise assisted data analysis (NADA) method is proposed, the ensemble empirical mode decomposition (EEMD). This new approach is based on the recent studies of the statistical properties of Fractional Gaussian noise (a versatile model for broadband noise include white noise proposed by Flandrin, 2004, and Wu and Huang, 2004), which showed that the EMD is effectively an adaptive dyadic filter bank when applied to fractional Gaussian noise [2].

III.1 EEMD Algorithm

The steps for EEMD are as follows [3].

- a) Initialize the number of ensemble I .
- b) Generate $x^i[t] = x[t] + w^i[t]$ ($i = 1, \dots, I$) are different realization of white Gaussian noise.
- c) Each $x^i[t]$ ($i = 1, \dots, I$), is fully decomposed by EMD getting their modes $IMF_k^i[t]$, where $k=1, 2, \dots, K$ indicates the modes.
- d) Assign \overline{IMF}_k as the k -th mode of $x[t]$, obtained as the average of the corresponding IMF_k^i : $\overline{IMF}_k[t] = \frac{1}{I} \sum_{i=1}^I IMF_k^i[t]$.

Just as the EMD method, the given signal, $x(t)$ can be reconstructed according to the following equation:

$$x(n) = \sum_{k=1}^K \overline{IMF}_k(t) + \bar{r}(t)$$

where $\overline{IMF}_k[t] = \frac{1}{I} \sum_{i=1}^I IMF_k^i[t]$ and $\bar{r}(t) = \frac{1}{I} \sum_{i=1}^I r_i(t)$

The EEMD described here employs all the important characteristics of noise. Its principle is simple: when a collection of white noise is added to the target signal it cancels each other out in a time space ensemble

mean. The reason is obvious that the added white noise would populate the whole time-frequency space uniformly with the constituting components of different scales separated by the filter bank.

IV. EXPERIMENT AND RESULT

Electrocardiogram (ECG) and non-stationary signal is analyzed in the present paper. Here the data length of ECG signal is 983 samples. In fig. 2 the added noise is only white Gaussian noise and reconstructed signal is shown in fig. 3. Here other than white Gaussian is also added to ECG signal to check the performance of EEMD algorithm. Three different patterns of added noise are [4]:

- a. EMG noise: Electromagnetic sources from the environment may overlay or cancel the signal being recorded from a muscle. It can be modeled by a random number with normal distribution, originally manipulated with the Matlab code `randn.m`. The maximum noise level is $(1/8)$ V.
- b. Power line interference: Power line interference is modelled by 50 Hz sinusoidal function with multiplication on amplitude derived with Matlab code `rand.m`. The maximum noise level is $(1/4)$ V.
- c. Baseline wander: Baseline wander is modelled by a Baseline wander a 0.333 Hz sinusoidal function. The maximum noise level is $(1/8)$ V.

The reconstructed signal via EEMD method adding above noise is shown in figure (5). Another signal having data length 500 samples is shown in figure (6) and its reconstructed signal is shown in figure (8).

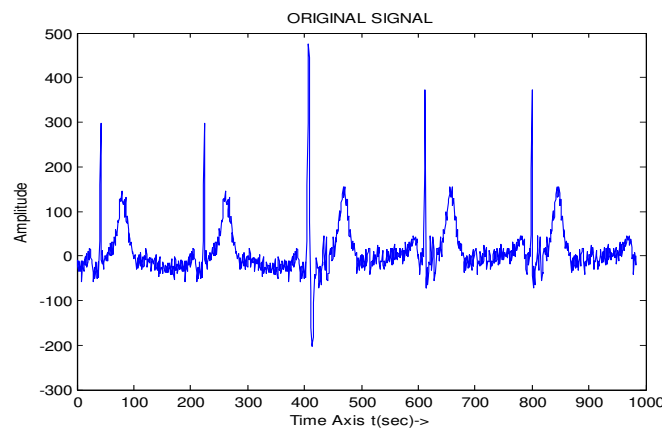


Fig.(1): ECG signal with data length 983 sample data

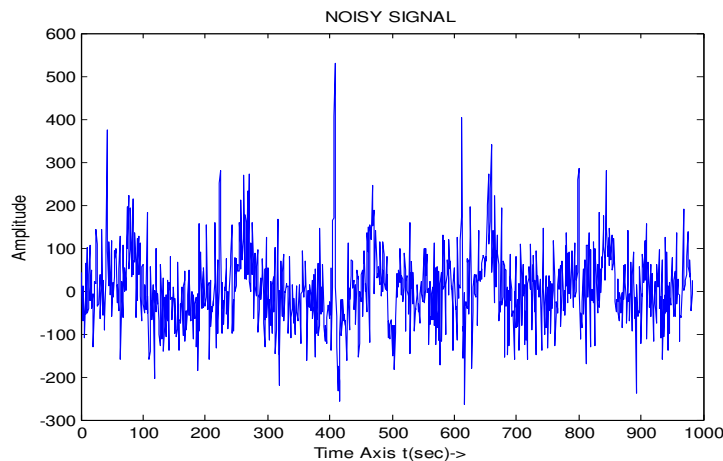


Fig.(2): ECG signal with White Gaussian noise (SNR= -1dB)

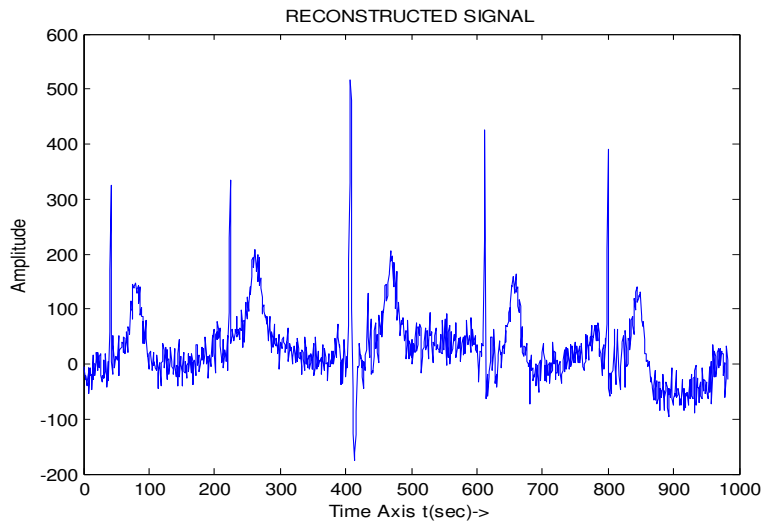


Fig.(3): Reconstructed ECG signal via EEMD having I=10

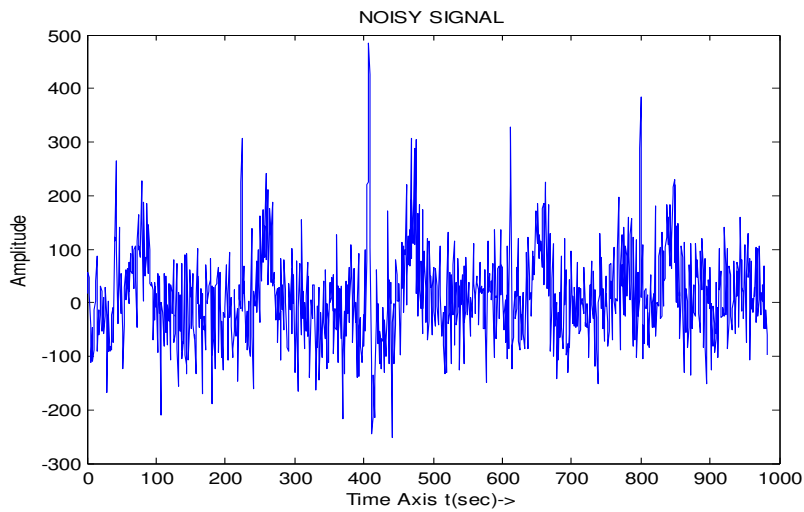


Fig.(4): ECG signal with artifacts

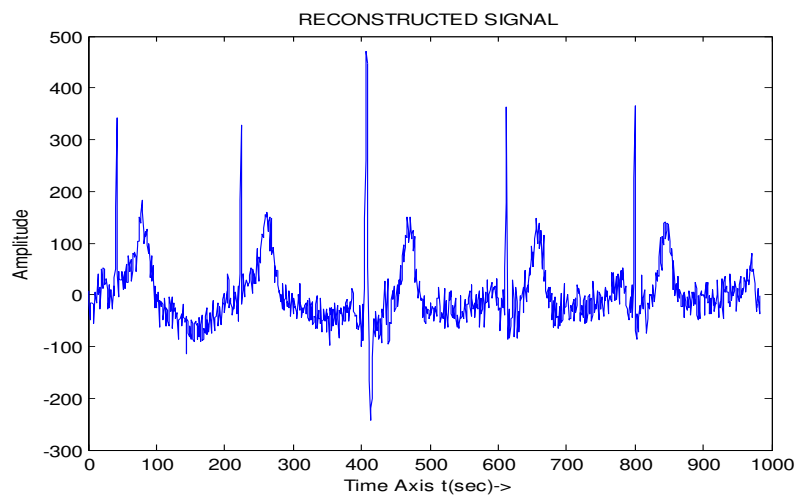


Fig.(5): Reconstructed ECG signal via EEMD

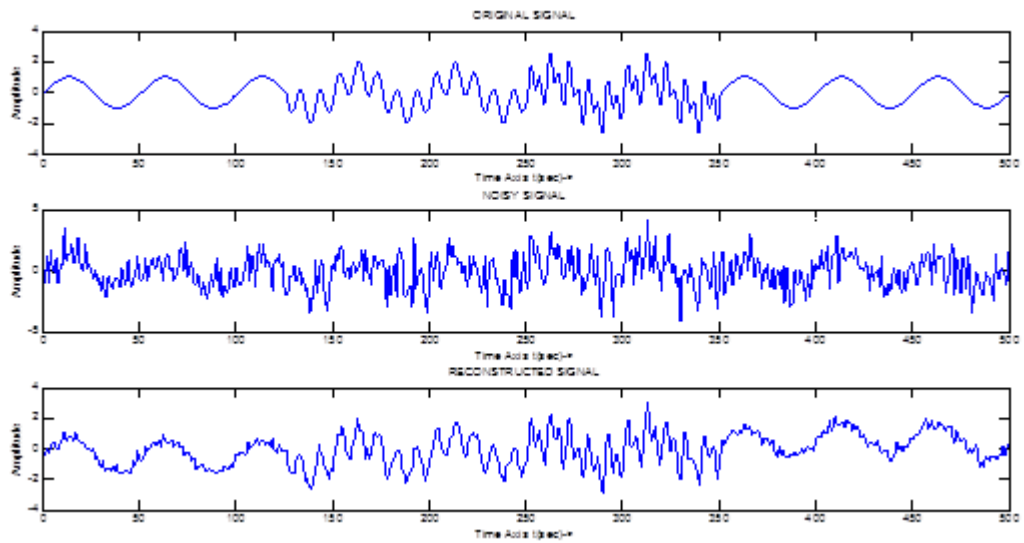


Fig.(6): Non-stationary Signal decomposed by EEMD having SNR= -1 and I=20

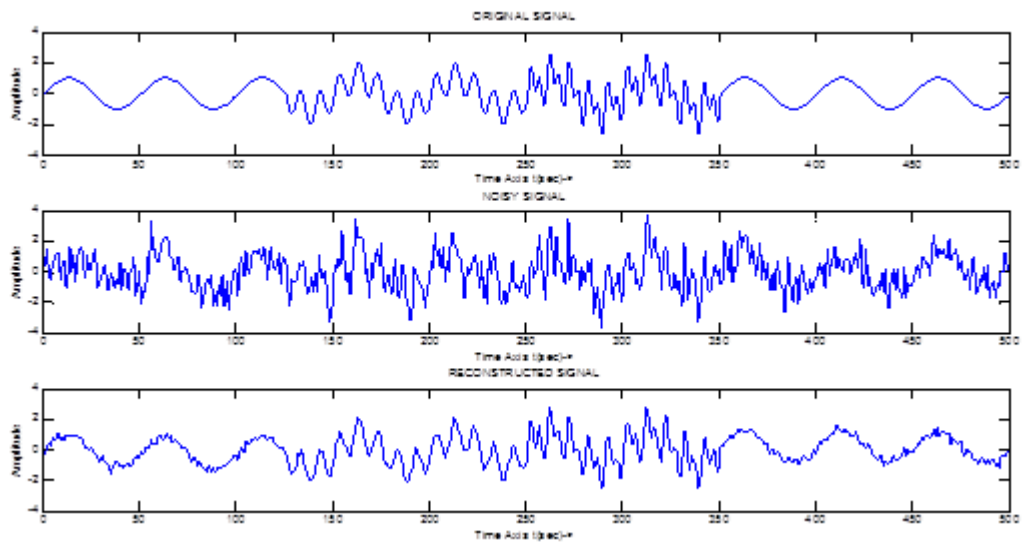


Fig.(7): Non-stationary Signal decomposed by EEMD having SNR= 1 and I=20

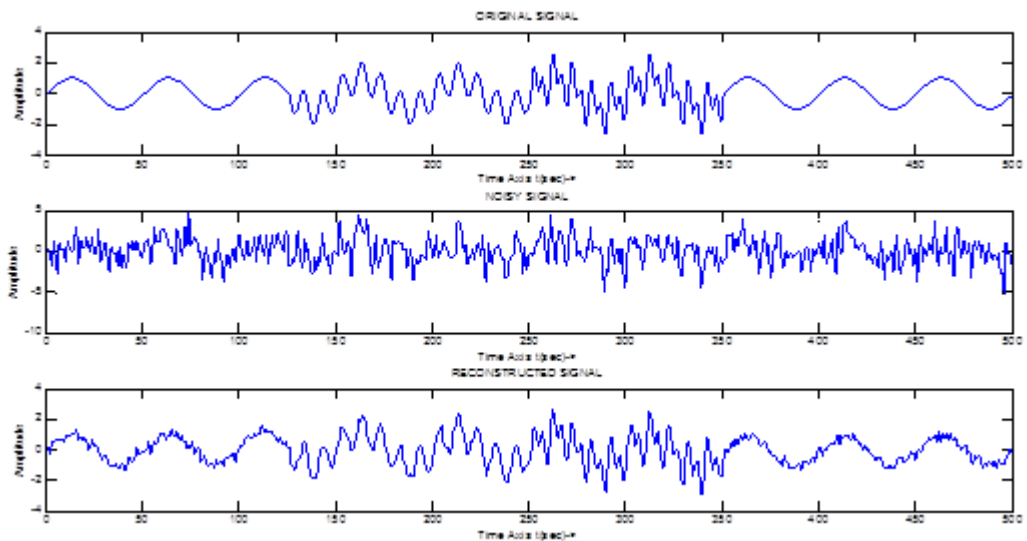


Fig.(8): Non-stationary Signal decomposed by EEMD having SNR= -3 and I=40

V. CONCLUSION

A new technique for the signal analysing and denoising has been described in this paper. Simulations results of the synthesized signals have expressed the effectiveness of the new algorithm. The technique differs from many conventional and EMD based algorithms as it uses noise to analysis signals and noise reduction. The method is a fully data-driven approach and the described technique has the ability to reduce noise efficiently for a large class of signals including almost all real and non-stationary signals.

REFERENCES

Proceedings and conference papers:

- [1]. N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu. ‘The Empirical Mode Decomposition and Hilbert Spectrum for Non-linear and Non stationary Time Series Analysis’. Proceedings of the Royal Society London A., 1998 454,903–995
- [2]. P. Flandrin, G. Rilling and P. Goncalves, ‘Empirical mode decomposition as a filter bank’, IEEE Signal Process. Lett. 11 (2004) 112–114.
- [3]. Z. Wu and N. E. Huang, ‘Ensemble empirical mode decomposition: a noise-assisted data analysis method,’ Advances in Adaptive Data Analysis, vol. 1, no. 1, pp. 1–41, 2009

Journal Papers:

- [4]. Kang-Ming Chang ‘Arrhythmia ECG Noise Reduction by Ensemble Empirical Mode Decomposition’ *Sensors* 2010, 10, 6063-6080; doi:10.3390/s100606063