An Improved Empirical Mode Decomposition for Long Signals

J.L. Sánchez*, Manuel D. Ortigueira†, Raul T. Rato‡, and Juan J. Trujillo§
* Departamento de Ingeniería Informática y de sistemas
Universidad de La Laguna 38271 La Laguna, Tenerife, Spain
Email: jsanrosa@ull.edu.es
† UNINOVA and DEE of Faculdade de Ciências e Tecnologia da UNL
Caparica,Portugal
Email: mdo@fct.unl.pt
‡ UNINOVA and Escola Superior de Tecnologia do Instituto Politecnico de Setubal,
2910-761 Setubal, Portugal
Email: raul.rato@estsetubal.ips.pt
§ Departamento de Análisis Matemático,
Universidad de La Laguna 38271 La Laguna, Tenerife, Spain
Email:jtrujill@ullmat.es

Abstract—**The analysis of long signals is relevant in many fields, as biomedical signal analysis. In this paper, a revision of the Empirical Mode Decomposition (EMD), from the application point of view, is done. The increase in number of Intrinsic Mode Functions (IMF) and computational time in long signals are the main problems that have been faced in this work. A solution based on a sliding window is proposed. An adaptive process is used to calculate the size of the sliding Windows. As a result, the effectiveness of the proposed algorithm increases with the length of the signal. Two examples are introduced to illustrate both problems mentioned above.**

Keywords–Empirical mode decomposition; Intrinsic mode function; long signals

I. Introduction

The Empirical Mode Decomposition (EMD), as was proposed initially by Huang et al [1] is a signal decomposition algorithm based on a successive removal of elemental signals: the Intrinsic Mode Functions (IMF). These are continuous functions such that at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. They are obtained through an iterative procedure called sifting that is a way of removing the dissymmetry between the upper and lower envelopes in order to transform the original signal into an amplitude modulated (AM) signal. Moreover, as the instantaneous frequency can change from instant to instant, it can be said that each IMF is a simultaneously amplitude and frequency modulated signal (AM/FM). So, the EMD is nothing else than a decomposition into a set of AM/FM modulated signals [2]–[5].

It must be emphasized that EMD is merely a computational algorithm that expresses a given signal as a sum of simpler components. It can not be said that the obtained components are true parts of the signal at hand.

The original algorithm had some implicit difficulties [3], [6]: extrema location, the end effect and the stopping criterion are critical. Some solutions were proposed in [3] and implemented in an algorithm that can be found at [7]. The location and amplitude of the extrema were estimated using a parabolic interpolation. To render less severe the end effect, the maxima and the minima were extrapolated by both sides. A new stopping criterion in the sifting procedure by introducing two resolution factors was defined.

In the last years, several modifications have been proposed to increase the performances of EMD, [8]–[16], [18], [19]. It is important to question if the introduced complexity compensates the quality increase. In this work, it will be preserved the simplicity of the original algorithm while it is increased the reliability and applicability of the decomposition.

In practical applications, there are several tradeoffs among resolution, signal length, the number of IMFs and running time. In fact, an increase in signal length produces two unwanted side-effects. On the one hand, it leads generally to a corresponding increase in the number of IMFs. Consequently, the running time may become so high, that the algorithm will be useless. The increase in the number of IMFs is a very important drawback because it may originate “false” components that are added to one IMF and subtracted to another one or appear isolated. So, in general, there are no guarantees to have IMFs that are really present in the original signal. On the other hand analyzing long signals with EMD is time-consuming or even impossible in a reasonable time [20] due to the fact, that spline interpolation of a large number of points takes a lot of computer resources. In applications to long signals [17], [21] the number of components and the running time would be so high that the algorithm would be almost useless. This problem was recently considered in [20] where the need of a more efficient and faster algorithm to deal with long signals was stated.

The algorithm described in [3] (and other similar EMD algorithms) is not prepared to deal with long signals as it uses only one window with the length of the signal. In section III it is shown the increase in processing time with increasing lengths. In some applications like EEG or ECG procesing we may have to process signals with lengths above 10^6. The processing time makes the existing algorithms almost useless. This suggests it is important to have an algorithm with the same characteristics but faster. In order to reach such goal, a sliding window EMD is proposed where consecutive windows overlap over a pre-specified amount. While [20] proposes to obtain a full EMD at every window, this approach can cause errors when dealing with signals that have fast changes in their

 ISBN: 978-1-61208-487-9

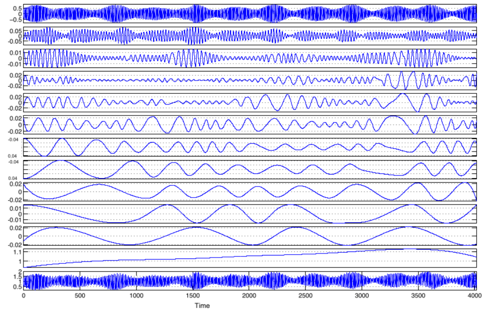

Figure 1. EMD of a tidal signal (in the last strip).

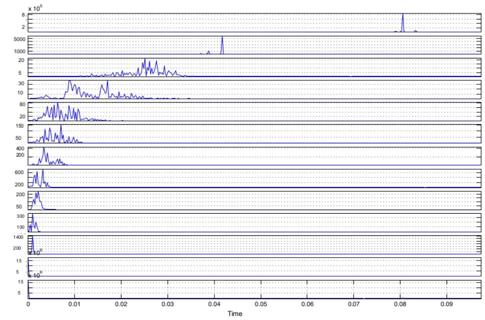

Figure 2. EMD of a tidal signal (in the last strip).

frequency composition. In such case, it could be impossible to obtain a decomposition with the same number of IMF's in all the segments. Our solution follows a different approach. Each IMF is calculated in sliding windows to ensure the number of obtained IMF's components is the same for the whole length of the signal. While [20] proposes to increase the size of the sliding window by a fixed quantity, here it is proposed to duplicate the size of the window when necessary. In that manner, fewer steps are involved for very long signals. Such implies that the length of the signal must be a power of two of the initial window size. Finally, a sliding window algorithm has an overload in the computation of the EMD for signals with a small number of samples. So, the minimum window size must have a lower threshold, corresponding to the length for which the sliding window EMD is slower than the whole length EMD. All these factors together led to an adaptive algorithm. Both, the length of the sliding windows and the length of the overlapping region depends on the length of the whole signal. Also, the overlapping region is tapered to improve the juction between adjacent windows.

The paper outlines as follows. In Section II, some reflections about the EMD are done. In Section III a new method, to analyse long signals, based on a sliding window is proposed. In section IV some illustrating results of the new method are presented. An application to the fan heater example and a comparison with the method in [3] is presented.

II. Some reflections about EMD

It is important to refer the usefulness of EMD in practical applications. A large number of papers published in the last years confirm the affirmation. One of the most important advantages of EMD is the ability to decompose a complex signal into a finite set of narrowband signals without introducing any particular constraint on its characteristics. This makes easier the spectral estimation and creation of simple models.

- Meaning of the IMFs
  In general, it is not possible to establish any special connection between a given IMF and the structure (eventually tied with the underlying physics) of the original signal. This does not mean that it can not be done in some particular situations, as in the case illustrated in Figure 1 where a tidal signal and its EMD are shown.

  A close look seems to point out that the most important IMFs are the two upper ones. The Fourier transform confirms such assumption since the peak frequencies of such IMFs correspond to the frequencies of the main components in the tidal signal: the positions of the Moon and the Sun relative to Earth and the Earth's rotation. The first has a period of about 12 hours and 25 minutes and the second has a period of 24 hours.

  These are clearly identified in the pictures. Even with a careful study it would be more difficult to give some meaning to some of the other components.

- The existence of false components in the IMF's.
  The above example calls the attention to the existence of false components. This can be seen, for instance, in doing a comparison of strips 3 and 4 in Figure 2 where a very similar spectra can be observed. This is a consequence of the numerical errors in sifting: one component is added in one IMF and subtracted in another one.

- The number of IMFs depends on the length of the signal.
  In fact, the number of components increases with the length of the signal. This may be an unwanted feature of the algorithm that is connected with the false component generation. On the other hand, this brings another drawback: the increase in the time required to do the decomposition.

An example

In a search for long range processes, an experiment with the electric circuit of a heater fan was carried out. The signal was sampled during two hours with a sampling interval of 10 ms. With it, the EMD for increasing length segments, using the algorithm described in [3], is computed and the results for 2 different resolutions (45 and 50) are shown in Table 1. Computations were carried out on a PC using MatLab. It is possible to get some decrease in the computational time by implementing the algorithm with a high-level language like $C\#$.

It can be concluded that the main drawbacks of EMD are the false components and the large computational time when the signal is long. In the following, we will propose a solution for the second problem that alleviates the first.

TABLE I. IMFs AND COMPUTATIONAL TIME FOR A HEATER FAN SIGNAL

Resol	length	IMFs	time (seconds)
45	11400	14	28
45	27600	15	41
45	114000	18	363
45	340800	20	1057
45	691800	22	3075
50	11400	13	29
50	27600	16	69
50	114000	19	602
50	340800	22	1774
50	691800	22	5195

III. DECOMPOSING LONG SIGNALS

As referred above, the objective of this paper is to propose an algorithm that can be used with long signals with significative reduction in the processing time and eventually in the number of IMFs.

A. The problem

Let $x(t)$ be a given signal to be decomposed by EMD. As referred above the number of IMFs is not known in advance and normally grows up with increasing the length of $x(t)$. This increments the computational burden, leading in some situations to very large computational times making the algorithm useless unless suitable actions are developed – see Table 1. One obvious procedure is to cut the signal into segments. However, this can lead to poor results due to the following

- Different number of IMFs from segment to segment;
- The end effects introduce discontinuities at the junction points.
- Reduced number of extremes leading to poor envelopes.

In [20], there is an attempt to overcome these problems by applying the traditional EMD algorithm to segments of the signal with a fixed number of IMFs. Although this algorithm reduces the computational time, it does not perform a complete true EMD.

In this paper, an algorithm suitable for obtaining the IMFs of very long signals is presented. This situation is very common in mechanical, electrical, and biomedical signal processing [22].

B. The solution

As it has been indicated one of the drawbacks in analyzing long signals is the computational load. The spline interpolation of a large number of points takes a lot of computer resources [20].

For this reason, the use of the sliding window EMD is proposed. The underlying idea to all the algorithms that use sliding windows is to use a divide and conquer strategy. The computation time for spline interpolation is dramatically reduced using smaller segments. However, due to the above referred constraints, too short segments can not be used. It must be taken into account that a sliding window algorithm with few points in each window has an overload due to the repetition of the interpolation. Depending on the features of the computer system used, there exist a threshold below which the classic algorithm is faster than the sliding window method. So, using short windows will result in an increase of the running time. The initial window length is calculated taking into account two factors: The first is that the signal length must be a power of two of the initial window size. The second is the threshold mentioned above. This threshold is set by the user as it depends on its computer system features.

The main idea of the algorithm is to apply the EMD sifting segment to segment to obtain only one IMF at a time. This procedure is done along the whole signal. This ensures that a real EMD is obtained. A pseudocode of our algorithm is:

Input
  $Filename$
  $StartingSample$
  $SignalLength$
$Resolution(dB)$
$OverlapPercentage$
$MinWindowSize$
$MinOscStop$
$MinOscEndStop$
START
  $OptimumSizeWindow = f(SignalLength)$
  $OverlapPoints = f(MinWindowSize$
$, OverlapPercentage)$
  $OSC \leftarrow Inf$
$MINOSC \leftarrow MinOscStop$
WHILE $(OSC > MinOsc)$
  IF $windowsize < Length$

$MINOSC = MinOscEndStop$

$SWEMD(1 : Length, L0, MINOSC)$
  ELSE
$MINOSC = MinOscEndStop$
$EMD(L, MINOSC)$

Figure 3. Adaptative Sliding Window EMD

For a general formulation consider a signal of length L. Select the segment length N and the overlapping M points.

1) Determine the starting window size, the number of samples in the overlap region, and the number of residual samples that do not fit in an integer number of windows.
2) Start a loop to obtain the whole set of IMF.
    a) Start a loop to obtain an IMF on a sliding window basis.
    b) The first window size determines the number of iterations of the sifting process for the rest of the IMF. The stopping criteria for a given IMF is the resolution in dB as proposed in [3].
    c) The process stops when the last segment is processed and the whole IMF is obtained
    d) Continue obtaining IMF's until the stopping criteria for small windows is reached and duplicate the window size.
3) Once the window size equals the whole length, the process continues with a fixed size until the number

of obtained extrema is less or equal than two.

4) Obtain the residual of the decomposition to have the whole EMD decomposition.

The main part of the algorithm is the outer While loop. It controls the stopping criteria for the EMD. The user must select how many oscillations are allowed in the last stage of the actual level. Once it is reached, the window size is duplicated. The process enlarges the size of the window adaptatively as it is needed to analyze components with bigger wavelengths. Moreover, enlarging the window size allow a better interpolation between distant extrema.

Each IMF is calculated with a sliding window if the window size is smaller than the whole signal length. This situation is evaluated in the first part of the IF statement. The situation in which there is only one window for the whole signal is evaluated in the ELSE part.

Segments must be tapered applying a complementary symmetrical window to avoid discontinuities at the boundaries. The windows are overlapped on both sides. A complementary symmetrical window is applied to consecutive segments. The function $\cos^2(\frac{\pi}{2M}n), \quad n = 0, 1, \ldots, M$, is used on the right of the segment and $\sin^2(\frac{\pi}{2M}n), \quad n = 0, 1, \ldots, M$, on the left. Of course, other windows can be used, as it is the case of the triangular.

To implement this process it is necessary to take into account the following observations:

1) The starting window size is correlated to the length of the signal to be analyzed. A simple possibility is making the starting segment a sub-power 2^{-k} of the signal length. On the other hand, a minimal window size must be imposed to avoid an excessive number of partitions. In this manner, the last window will cover the whole length signal. It must be taken into account that, dividing the signal length by powers of two can lead to a non-integer size of the starting window; the integer part of the quotient is used. So the signal length is covered by an integer number of windows and a residual. Depending on its size this residual can be assigned to the last window, enlarging its size or constitute a new window, usually with different size to the previous one.
2) Regarding the criterion to determine when to enlarge the signal, it must be taken into account the fact that the minimum frequency to be analyzed depends on the window size (as it has been indicated before). As the sifting process requires oscillatory signals, it must be ensured that the window contains at least a minimum number of periods that can be selected by the user. The criterion of duplicating the window size when the number of extrema in the previous IMF is lower than a user selected threshold was adopted.
3) Concerning the overlap region, it must be taken into account that it is necessary to reduce the undesirable end effects. Extrapolation is not used, since there are enough number of samples outside the actual segment. That implies that overlapping consecutive windows solve both, boundary and end effects.

IV. ILLUSTRATING RESULTS

In the following, the behavior of the algorithm is illustrated. Firstly, the fan signal mentioned above is decomposed using the algorithm in [3]. The result is shown in Figure 4. Secondly, the same signal is decomposed using the method proposed here. The result can be observed in Figure 5. A comparison of the computation time and the number of IMF obtained with both algorithms is presented in Table 2.

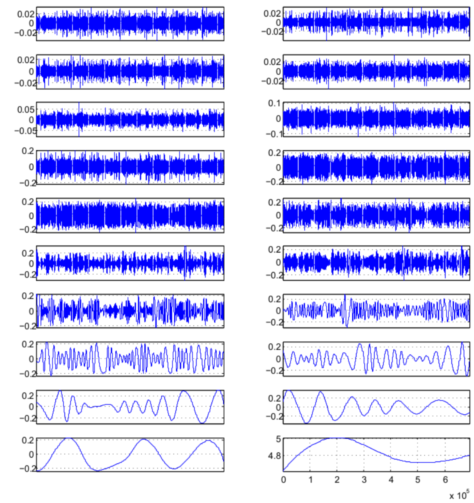

Figure 4. EMD using algorithm of [3]

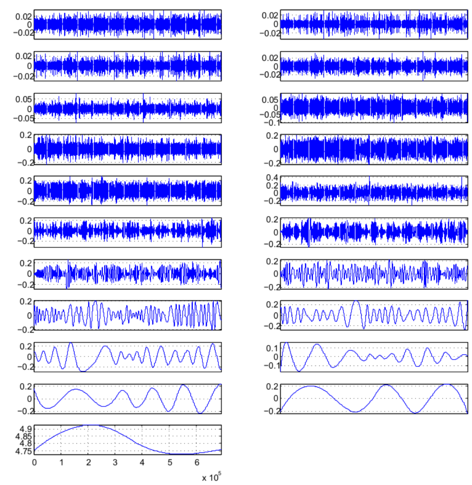

Figure 5. EMD using algorithm proposed here.

TABLE II. TIME COMPARISON BETWEEN BOTH ALGORITHMS FOR A HEATER FAN SIGNAL

Resol	length	IMFs	time EMD as in [3]	IMFs	time new EMD
45	11400	14	28	14	17
45	27600	15	41	14	37
45	114000	18	363	16	159
45	340800	20	1057	19	452
45	691800	22	3075	20	923
50	11400	13	29	12	23
50	27600	16	69	15	50
50	114000	19	602	18	217
50	340800	22	1774	20	669
50	691800	22	5195	21	1414

 ISBN: 978-1-61208-487-9

While for 11,400 samples the sliding window computation time is 79% of the EMD calculated as in [3], for 691,800 samples the computation time is only 27%. That is due to the fact that the window size increases by powers of two, which results in smaller running times for very long signals. So, the effectiveness of the proposed adaptive sliding window algorithm increases with the length of the signal. Despite of the fact of using many windows for the calculation, the obtained IMF's show a high quality as no discontinuities can be observed in the last IMF's for a signal with more than 600,000 samples.

It must be taken into account that any error in a given IMF is propagated to the rest of the decomposition. As subsequent components have smaller amplitudes, the errors have greater importance. Our algorithm has shown a good behavior as it can be observed in 5. That is due to the fact that averaging the overlapped region between two consecutive windows smooths the result. As the number of samples in the overlapping region is based on a fixed percentage of the window size, the number of samples change with the window size.

V. CONCLUSIONS

The Empirical Mode Decomposition is a technique to decompose any signal into a finite set of narrowband components, the Intrinsic Mode Functions. The number of components and computational time increase dramatically when the length of the signal becomes large. Proposals for solving this problem had been done, but without the required quality. A modified sifting algorithm to deal with long signals was proposed here. It is based on computing every IMF using a sliding window. The algorithm is adaptive as both, the length of the sliding windows and the overlapping region depends on the signal to be analized. The change on the length of the sliding window by powers of two has two positive consequences. On one hand, the final window will cover the whole length of the signal in a few steps. On the other hand, the effectiveness of the proposed method increases with the length of the signal. The application of our method confirms the affirmation.

ACKNOWLEDGMENTS

This work was partially funded by National Funds through the Foundation for Science and Technology of Portugal, under the project PEst-UID/EEA/00066/2013 and by project MTM2013-417 from Government of Spain.

REFERENCES

[1] Norden E. Huang et al., "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis." Proceedings of the Royal Society of London A, vol. 454, issue 1971, 1998, pp. 903–995.

[2] S. Peng, X. Hu, and W. L. Hwang, "Multicomponent am-fm signal separation and demodulation with null space pursuit." Signal, Image and Video Processing, vol. 7, issue 6, 2013, pp. 1093-1102.

[3] R. T. Rato, M. D. Ortigueira and A. G. Batista, "On the HHT, its problems, and some solutions." Mechanical Systems and Signal Processing, vol. 22, issue 6, 2008, pp. 1374–1394,

[4] R. T. Rato, M. D. Ortigueira and A. G. Batista, "The EMD and its use to identify system modes." In Proceedings of the International Workshop on New Trends in Science and Technology [CD-ROM], Nov. 03-04, 2008, Ankara, Turkey.

[5] M. K. Hasan, K. M. S. Apu and M. K. I. Molla, "A robust method for parameter estimation of ar systems using empirical mode decomposition." Signal, Image and Video Processing, vol. 4, issue 4, 2010, pp. 451-461.

[6] P. C. Chu, C. W. Fan and N. Huang, "Compact empirical mode decomposition: an algorithm to reduce mode mixing, end effect, and detrend uncertainty." Advances in Adaptive Data Analysis, vol. 4, issue 3, 2012, pp. 1250017 (18 pages).

[7] M. Ortigueira, "Empirical Mode Decomposition [online]" Available: http://www.mathworks.com/matlabcentral/fileexchange/21409-empirical-mode-decomposition. [Accessed: 18- Apr- 2016]

[8] K. M. Chang and S. H. Liu, "Gaussian noise filtering from ECG by wiener filter and ensemble empirical mode decomposition." Journal of Signal Processing Systems, Special Issue "Signal Processing Circuits and Systems for Bio-Signals", vol. 64, issue 2, 2011, pp. 249-264.

[9] M. A. Colominas, G. Schlotthauer and M. E. Torres, "Improved complete ensemble EMD: A suitable tool for biomedical signal processing." Biomedical Signal Processing and Control, vol. 14, 2014, pp. 19–29.

[10] M. Feldman, "Analytical basics of the EMD: Two harmonics decomposition." Mechanical Systems and Signal Processing, vol. 23, issue 7, 2009, pp. 2059-2071.

[11] X. Guanlei, W. Xiaotong, X. Xiaogang and Z. Lijia, "Improved EMD for the analysis of fm signals." Mechanical Systems and Signal Processing, vol. 33, 2012, pp. 181–196.

[12] H. Jiang, C. Li and H. Li, "An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis." Mechanical Systems and Signal Processing, vol. 36, issue 2, 2013, pp. 225-239.

[13] Z. K. Peng, P. W. Tse and F. L. Chu, "An improved Hilbert-Huang transform and its application in vibration signal analysis." Journal of Sound and Vibration, vol. 286, issues 1-2, 2005, pp. 187–205

[14] N. U. Rehman, C. Park, N. E. Huang and D. P. Mandic, "Emd via MEMD: Multivariate noise-aided computation of standard EMD." Advances in Adaptive Data Analysis, vol. 5, issue 2, 2013, pp. 1350007 (25 pages).

[15] P. Singh, P. K. Srivastava, R. K. Patney, S. D. Joshi and K. Saha, "Nonpolynomial spline based empirical mode decomposition." In International Conference on Signal Processing and Communication (ICSC), Noida, India, December 2013, pp. 435–440, IEEE.

[16] Y. Yang, J. Deng and D. Kang, "An improved empirical mode decomposition by using dyadic masking signals." Signal, Image and Video Processing, vol. 9, issue 6, 2013, pp. 1259–1263.

[17] F. Ebrahimia, S. K. Setarehdana and H. Nazeranb, "Automatic sleep staging by simultaneous analysis of ECG and respiratory signals in long epochs." Biomedical Signal Processing and Control, vol. 18, 2015, pp. 69–79.

[18] X. D. Yu, M. Y. Zhang, M. Q. Zhu, K. H. Xu and Q. C. Xiang, "An improved extension method of EMD based on svrm." Applied Mechanics and Materials, vol. 543-547, 2014, pp. 2697–2701.

[19] A. Eftekhar, C. Toumazou, E. M. Drakakis, "Empirical mode decomposition: Real-time implementation and applications." Journal of Signal Processing Systems, vol. 73, issue 1, 2013, pp. 43–58.

[20] P. Stepien, "Sliding window empirical mode decomposition – its performance and quality." EPJ Nonlinear Biomedical Physics, vol. 2, issue 1, 2014, pp. 2–14.

[21] Md. A. Kabir and C. Shahnaz, "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains." Biomedical Signal Processing and Control, vol. 7, issue 5, 2012, pp. 481–489.

[22] M. P. Pierzchalski, R. A. Stepien, P. Stepien, "New nonlinear methods of heart rate variability analysis in diagnostics of atrial fibrillation." International Journal of Biology and Biomedical Engineering, vol. 5, issue 4, pp. 201-208, 2011.