World Scientific
www.worldscientific.com

EMD OF GAUSSIAN WHITE NOISE: EFFECTS OF SIGNAL LENGTH AND SIFTING NUMBER ON THE STATISTICAL PROPERTIES OF INTRINSIC MODE FUNCTIONS

GASTÓN SCHLOTTHAUER[*,§], MARÍA EUGENIA TORRES[*,¶],
HUGO L. RUFINER[†,‖] and PATRICK FLANDRIN[‡,**]

*Lab. Signals and Nonlinear Dynamics, Applied Research group
on Signal Processing and Pattern Recognition (ARSiPRe)
Facultad de Ingeniería, Universidad Nacional de Entre Ríos
Ruta 11 Km 10 Oro Verde, Entre Ríos 3100, Argentina*

†*Lab.de Señales e INteligencia Computacional (SINC)
Fac. de Ingeniería y Cs Hídricas, Univ. Nac. del Litoral
CC 217, Ciudad Universitaria,
Paraje El Pozo, S3000 Santa Fe, Argentina*

‡*Laboratoire de Physique (UMR 5672 CNRS)
École normale supérieure de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, France*
§*gschlott@bioingenieria.edu.ar*
¶*metorres@santafe-conicet.gov.ar*
‖*lrufiner@fich.unl.edu.ar*
****patrick.flandrin@ens-lyon.fr*

This work presents a discussion on the probability density function of Intrinsic Mode Functions (IMFs) provided by the Empirical Mode Decomposition of Gaussian white noise, based on experimental simulations. The influence on the probability density functions of the data length and of the maximum allowed number of iterations is analyzed by means of kernel smoothing density estimations. The obtained results are confirmed by statistical normality tests indicating that the IMFs have non-Gaussian distributions. Our study also indicates that large data length and high number of iterations produce multimodal distributions in all modes.

Keywords: Empirical Mode Decomposition (EMD); Intrinsic Mode Function (IMF); Gaussian white noise; sifting.

1. Introduction

The Empirical Mode Decomposition (EMD), first proposed by Huang *et al.*, arose as a new completely data-driven method for signal analysis with many potential applications.[1] However, some aspects concerning its theoretical properties are still obscure and the results of its application on a given signal cannot be predicted. Studies have been carried out by Flandrin *et al.* which allow us to assert that, when

applied to Gaussian white noise, EMD acts in a similar way as a wavelet filter bank.[2,3] A research on the EMD effects on uniformly distributed white noise was performed by Wu and Huang, and normal distributions of the Intrinsic Mode Functions (IMFs) were reported.[4] On the basis of these results, Wu and Huang suggested a chi-square distribution for the energy density distribution of the IMFs. In the mentioned article, the authors used one million length white noise time series and the sifting process was limited to 10 iterations. A similar study was carried out by the same authors on Gaussian white noise, with 7–10 sifting iterations deriving similar conclusions.[5] However, in none of this works a normality test was applied and the conclusions were based on a Gaussian-like histogram estimation, using 50,000 data points in Refs. 4 and 5. An extension of these experiments to fractional Gaussian noise was conducted by Flandrin *et al.*, and a similar conclusion was reached about the normality of the IMFs.

In this work, we study the statistical properties of EMD on Gaussian distributed white noise. We show and discuss the dependence of the IMFs distribution on the signal lengths and the number of allowed iterations that involves the sifting process. Additionally, normality tests are carried out on the IMFs in order to test the null hypothesis of normality.

2. Experimental Design

In order to study the Probability Density Functions (PDFs) of the IMFs obtained by EMD of Gaussian white noise, a random signal with Gaussian distribution $x(n)$ was generated with zero mean and unitary variance, with 2^{20} data points. Then, the signal $x(n)$ was split into W non-overlapping windows $x^w(n)$ of lengths $L = 2^{10}, 2^{12}, \ldots, 2^{20}$. For each window $x^w(n)$, the EMD algorithm was applied using a given maximum number Ni of iterations in the sifting process ($Ni = 5, 10, 15, 25, 50$ and unlimited), yielding the $\text{IMF}_k^w(n)$ for modes $k = 1, \ldots, K$. In this way, we obtain a number of realizations to be used for the PDFs estimation purpose. Therefore, the here considered IMFs are built by concatenation:

$$\text{IMF}_k(n) = [\text{IMF}_k^1(n) \,|\, \text{IMF}_k^2(n) \,|\, \cdots \,|\, \text{IMF}_k^W(n)], \tag{1}$$

for $k = 1, 2, \ldots, K$.

We used the Matlab implementation of EMD available at Ref. 8. The sifting is ended when the number of zero-crossings and the number of extrema differ at most by one, and when the local mean between the upper and lower envelopes are close to zero. The criteria for deciding if such local mean is close to zero enough, are adopted as in Ref. 7, using the typical values suggested in this paper. If these criteria are not achieved, but an *a priori* established maximum number of sifting iterations is reached, the process is terminated.

In this way, using (1) for each IMF_k, the PDF was estimated (at the different window lengths L and number of iterations Ni) using $N = 2^{20}$ data points. The probability density estimation was performed by Gaussian kernel method,[9] using
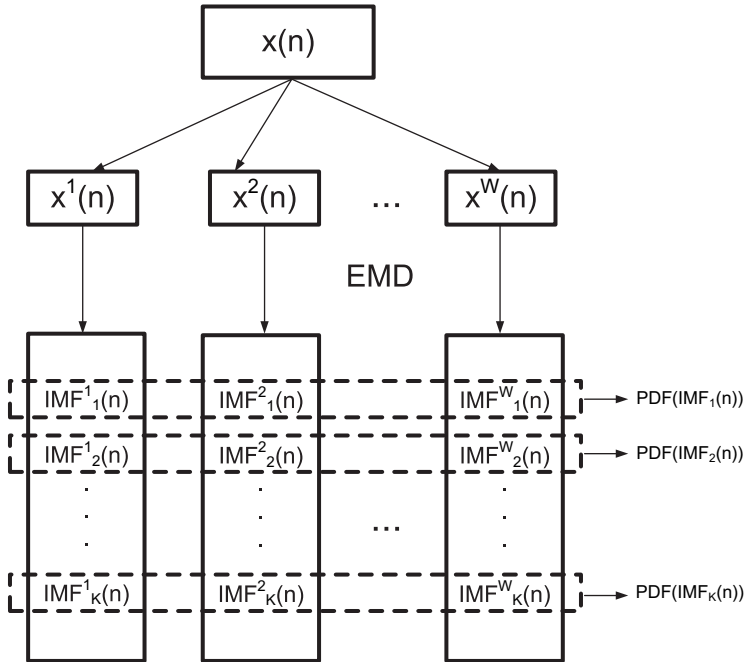
Fig. 1.   The signal $x(n)$ was split into nonoverlapping windows, $(x^1(n), x^2(n), \ldots, x^W(n))$. EMD was applied and the K modes $\mathrm{IMF}_k^w(n)$, $k = 1, \ldots, K$, were obtained for each $x^w(n)$, $w = 1, \ldots, W$. Next, the PDF was estimated for each $\mathrm{IMF}_k(n)$, $k = 1, \ldots, K$.

$N_s = 500$ equally spaced points y_s covering the range of data in the corresponding IMF_k:

$$\mathrm{PDF}(y_s) = \frac{1}{N} \sum_{n=1}^{N} \omega(y_s - \mathrm{IMF}_k(n); h), \tag{2}$$

where ω is the kernel function, whose variance is controlled by the parameter h. In our case, $\omega(z; h)$ denotes the normal density function in z with mean 0 and standard deviation h. For each mode k, the bandwidth h of the Gaussian kernel was chosen as the optimal for normal densities[9]:

$$h = \left(\frac{4}{3N}\right)^{1/5} \sigma, \tag{3}$$

where σ is its standard deviation.

The above described procedure is depicted in Fig. 1.

3. Results and Discussion

First we analyze and compare the PDFs corresponding to each of the IMF_k^w for modes $k = 1, 2, \ldots, 6$ for windows $x^w(n)$ of fixed length and unlimited sifting iterations (Sec. 3.1).

In order to discuss the incidence of the signal length and the maximum number of sifting iterations, we perform the analysis of the above described data PDFs using different approaches: (i) density estimation for a fixed value of window length (L) and different number of maximum iterations (Ni), (ii) density estimation for a fixed maximum number of sifting iterations (Ni) varying the data length (L). See Secs. 3.2 and 3.3, respectively.

The first case is equivalent to study the influence of data length on the statistical properties of the IMFs.

We also study the number of sifting iterations demanded by the method to accomplished the stopping criteria (Ni is unlimited) in relation to the data length L (Sec. 3.4).

In Sec. 3.5 a normality test is applied to the IMFs.

3.1. *IMF density estimation*

In Fig. 2, we compare the kernel smoothing density estimations of IMF_k^w, with $w = 1, 2, \ldots, 1024$ and $k = 1, 2, \ldots, 6$ for a Gaussian noise realization $x^w(n)$ of length $L = 2^{10}$ and Ni set to unlimited, plotted in gray lines. It must be noticed that in this way we are estimating the PDFs corresponding to $2^{10} = 1024$ realizations. The corresponding averages are displayed in black lines. The PDFs have been estimating using 50 Gaussian kernels.

It can be appreciated that the individual PDFs do not seems to be normal. However, the average of these estimated distributions tends to a normal curve. An exception is observed in the first mode, where a bimodal behavior can be appreciated. An explanation of this situation can be read in Ref. 3.

3.2. *Fixed data length*

By means of EMD we obtain the IMF_k for $k = 1, 2, \ldots, 10$ with $L = 2^{10}$, using different Ni values, ($Ni = 5, 10, 15, 20, 25, 50, \ldots, \infty$). The corresponding PDFs are estimated with a Gaussian kernel smoothing method using 500 equally spaced points that cover the range of each IMF_k. The obtained results for modes $k = 1, \ldots, 6$ are shown in Fig. 3. We can observe that for Ni larger than 5, the distributions look very similar within each mode. In every case, except the first mode, we can appreciate slightly spiky shape around the maximum that corresponds to the statistical mode.

A similar experiment is carried out with $L = 2^{20}$. In Fig. 4 are displayed the PDFs estimations of IMFs 1–6. In this case the PDFs exhibit a greater dependence on the Ni. It can be observed that, in all the cases, increasing Ni the PDF evolves from a "spiky" shape (alike those obtained with $L = 2^{10}$) towards a flattened one. In occasions bimodal ($k = 5$ and 6) and tri-modal ($k = 2, 3$ and 4) PDFs are observed when Ni is unlimited.

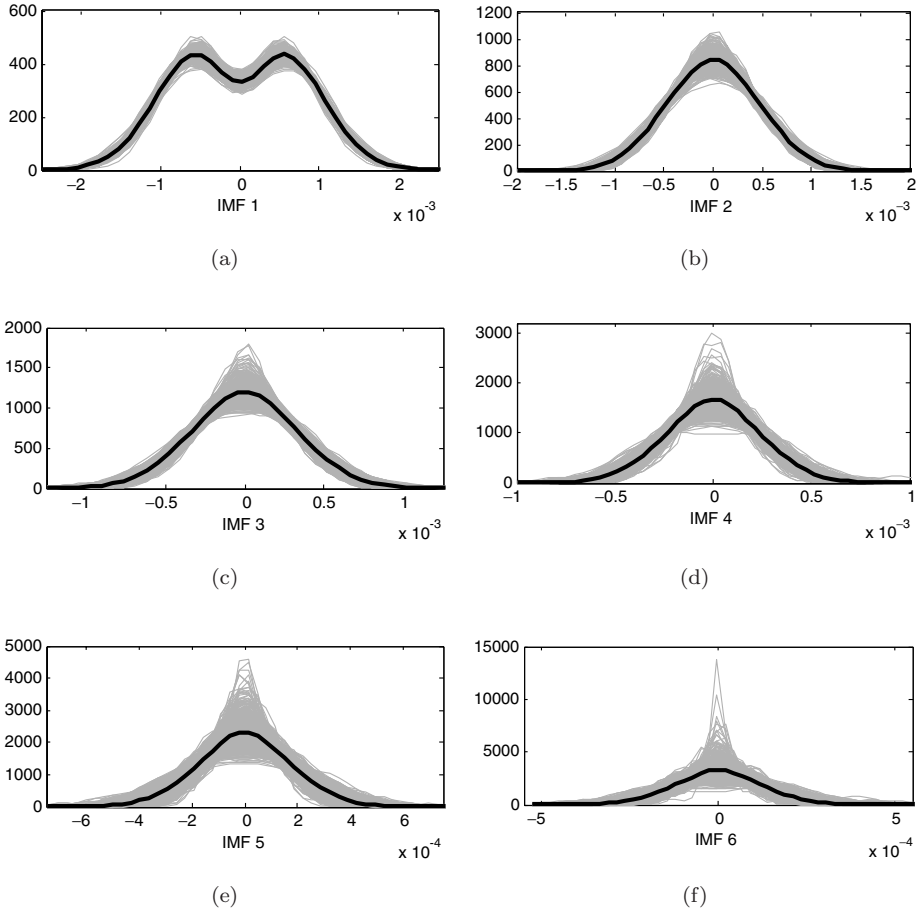These observations confirm to the need of performing a normality test.

Fig. 2. Kernel smoothing density estimation of IMFs 1–6 for $L = 2^{10}$ and unlimited sifting iterations. The PDF estimations for each $x^w(n)$, $w = 1, 2, \ldots, 1024$ are plotted in gray. The averaged PDFs are displayed in black. The averaged PDFs for IMFs 2–6 are approximately Gaussian.

3.3. *Fixed maximum number of iterations*

Setting $Ni = 5$, we obtain the PDFs shown in Fig. 5, depending on the data length $L = 2^{10}, 2^{12}, \ldots, 2^{20}$. Once more, the PDFs are very close within each mode, and seem not to depend on L. Their shapes look to be closer to a Laplacian distribution than to a Gaussian.

On the other hand, allowing an unlimited number of sifting iterations Ni, we observe in Fig. 6 that increasing the signal length L, the densities maxima go lower and the shape of the PDFs changes from unimodal to multimodal.

3.4. *Maximum number of iterations*

From a practical point of view, an interesting question about the previous experiments should be which is the number of iterations demanded by EMD to accomplish

(a) (b) (c) (d) (e) (f)

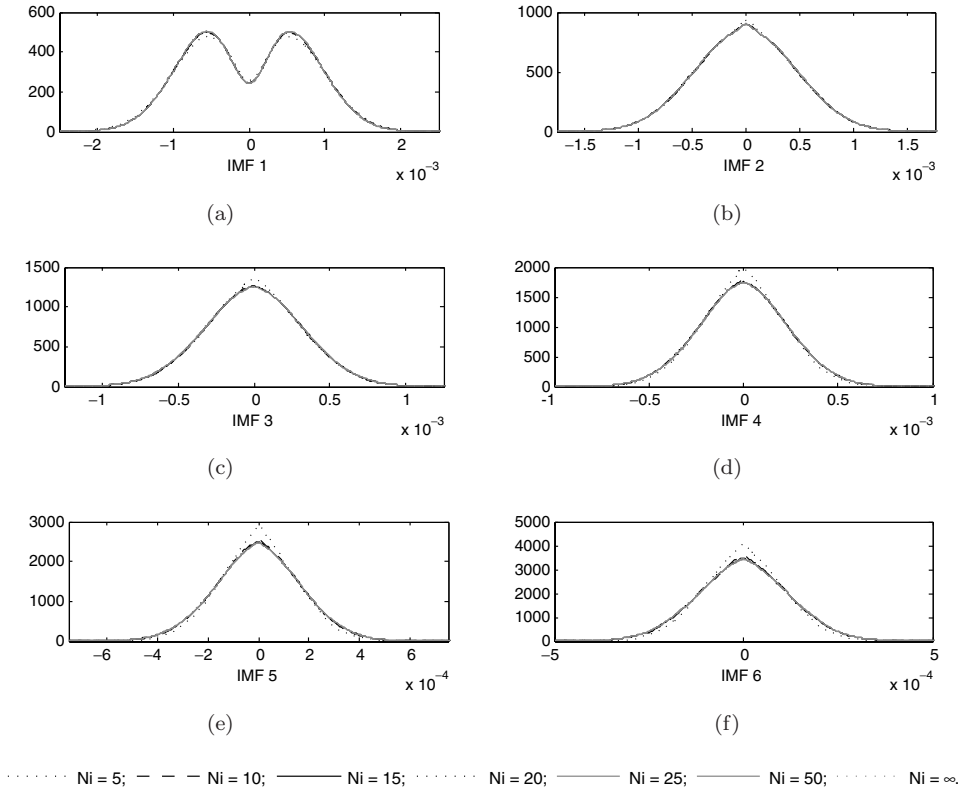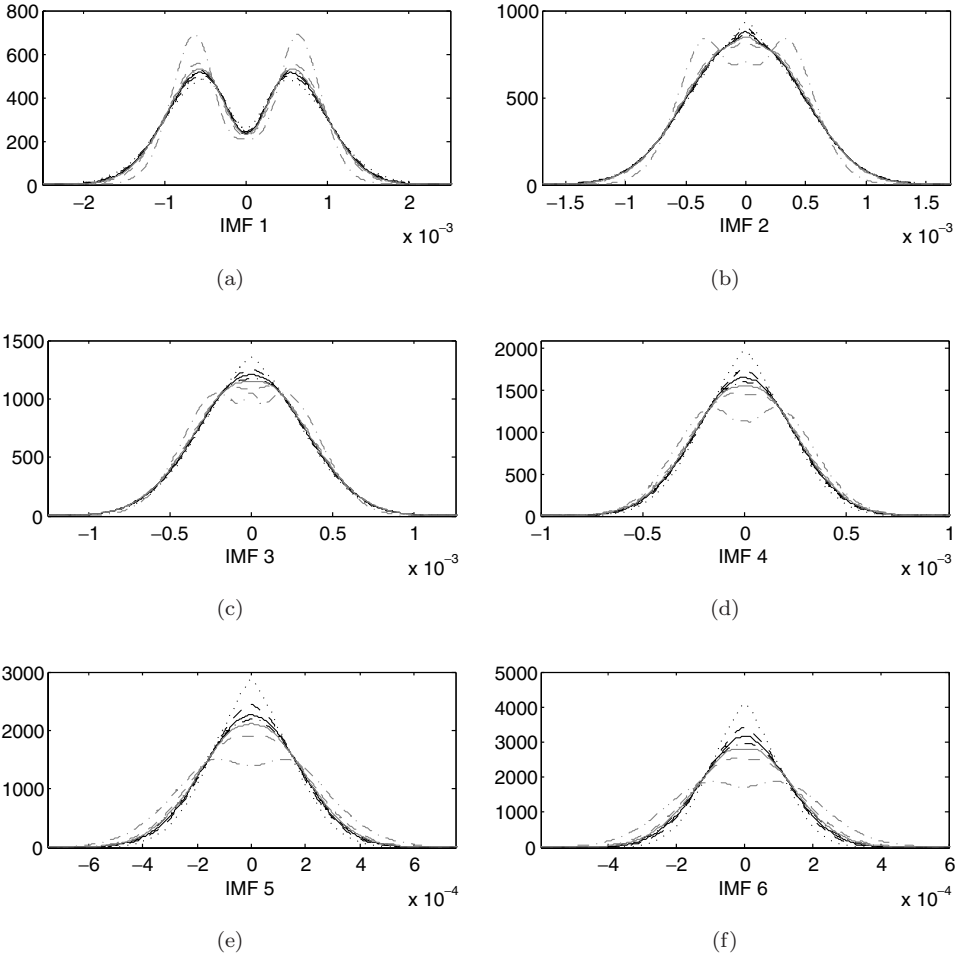Ni = 5; — — — Ni = 10; ——— Ni = 15; ⋯⋯⋯ Ni = 20; ——— Ni = 25; ——— Ni = 50; ⋯⋯⋯ Ni = ∞.

Fig. 3.   Kernel smoothing density estimation of IMFs. The PDF estimations of IMFs 1–6 are shown in (a)–(f). The IMFs are constructed by concatenation of the obtained IMFs of $x^w(n)$ with length $L = 2^{10}$. The maximum number of iterations varies: $Ni = 5, 10, 15, 20, 25, 50$ and unlimited.

the stopping criteria depending on the signal length. In order to answer it, we record this information and the results are shown in Fig. 7 for data length $L = 2^{10}, 2^{12}, \ldots, 2^{20}$. As L increases, then the median and the dispersion also increase.

When $L = 2^{10}$, we have 1024 realizations, therefore the outliers showed in the boxplots can be disregarded. The number of iterations needed to accomplish the stopping criteria has a median value of approximately 10 for all modes. This situation provides a possible explanation to the similarities of the PDFs presented in Fig. 3 for $Ni \leq 10$. However, when L increases, the number of needed iterations and their dispersion also increases.

3.5. *Normality test*

We have run the Jarque–Bera normality test[6] on the IMFs obtained as described in Eq. (1), applying EMD to Gaussian noise of lengths $L = 2^{10}$ and $L = 2^{20}$ for several maximum number of iterations ($Ni = 5, 10, 15, 20, 25, 50, \ldots, \infty$).

Fig. 4.   Kernel smoothing density estimation of IMFs. The PDF estimations of IMFs 1–6 are shown in (a)–(f). The IMFs are obtained by EMD of $x(n)$ with length $L = 2^{20}$. The maximum number of iterations varies: $Ni = 5, 10, 15, 20, 25, 50, \ldots, \infty$.

In the case of $L = 2^{10}$ the null hypothesis of normality has been always rejected with $\alpha = 0.05$ and $p < 0.001$. On the other hand, for $L = 2^{20}$ the null hypothesis has not been rejected only for IMF 6 when $Ni = 10$ ($p = 0.496$), IMFs 3, 4 and 5 when $Ni = 15$ ($p = 0.448$, $p = 0.500$, and $p = 0.378$, respectively), and IMF 6 when $Ni = 25$ ($p = 0.500$). The null hypothesis has been rejected with a larger value of p ($p = 0.002$) for IMF 4, corresponding to $Ni = 10$. Similar results were obtained by applying Lilliefors test of normality.

Looking at Fig. 3, the curves corresponding to those cases when the null hypothesis was not rejected seemed to be Gaussians. The occurrence of these Gaussian-like
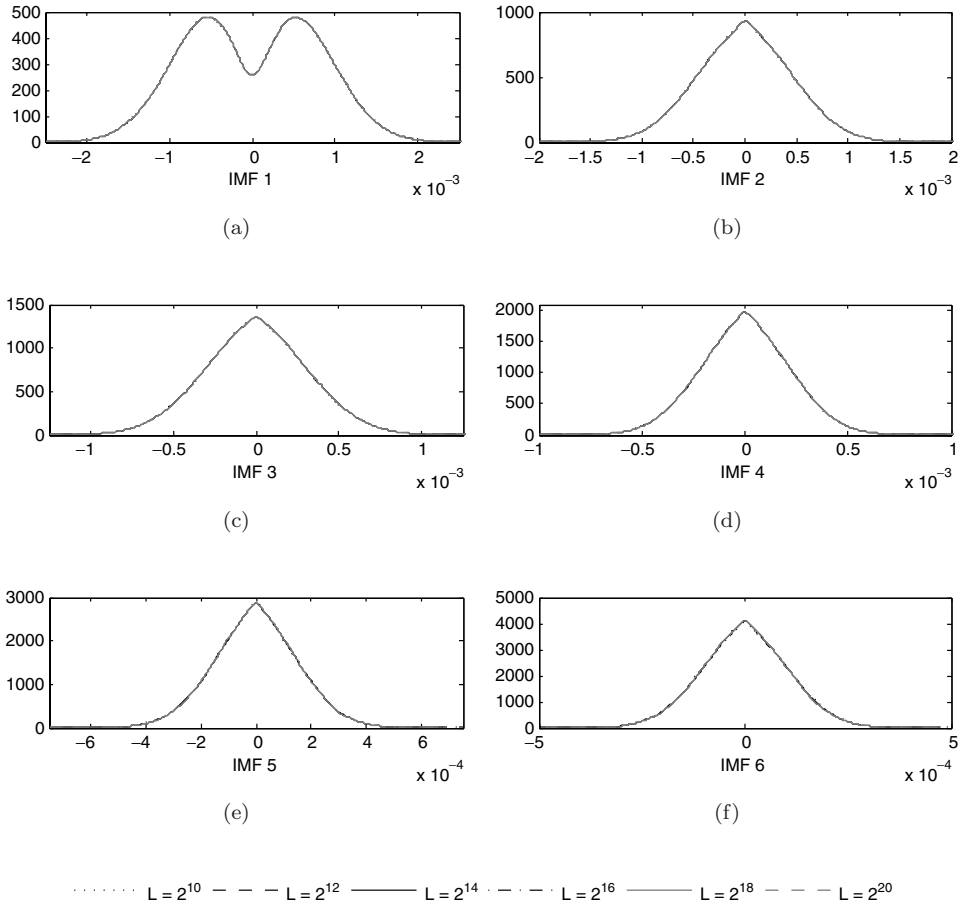
Fig. 5.   Kernel smoothing density estimation of IMFs. The PDF estimations of IMFs 1–6 are shown in (a)–(f). The IMFs are constructed by concatenation of the obtained IMFs of $x^w(n)$ with lengths $L = 2^{10}, 2^{12}, \ldots, 2^{20}$. The maximum number of iterations was fixed to : $Ni = 5$.

distributions, become evident while increasing Ni the curves change from "spiky" to multimodal alike shapes.

3.6. *Further remarks*

As discussed in the previous sections, a finding of this paper is that the shape of the PDFs of the IMFs have a tendency to become bimodal, or even multimodal, as the sifting number or the data length becomes large. Also, the first IMF always shows a bimodal structure. In order to explore this situation further and try to understand the mechanism behind this tendency, one could think in running similar experiments using, instead of the EMD, the ensemble empirical mode decomposition (EEMD) proposed by Wu and Huang, which overcomes some of the EMD modes mixing
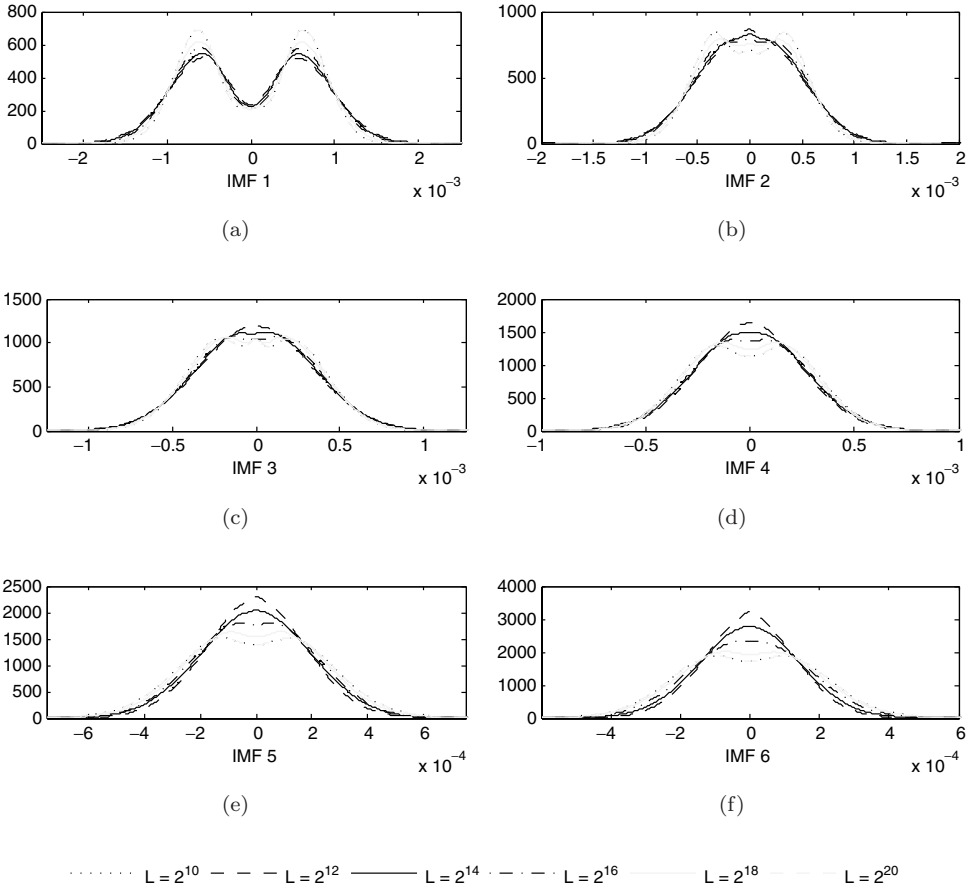
Fig. 6. Kernel smoothing density estimation of IMFs. The PDF estimations of IMFs 1–6 are shown in (a)–(f). The IMFs are constructed by concatenation of the obtained IMFs of $x^w(n)$ with lengths $L = 2^{10}, 2^{12}, \ldots, 2^{20}$. The maximum number of iterations was unlimited.

problems.[10] It defines the true IMF components as the mean of certain ensemble of trials of size N_e, each obtained by adding white noise of finite variance to the original signal.

In preliminary experiments using EEMD with only five sifting iterations and $N_e = 500$, the Gaussian null hypothesis could not be rejected in any mode except the first one, where a bimodal PDF was still present. In fact, this could be an expected result, because of the central limit theorem, given that at each mode the final IMF component is obtained by the addition of the mode corresponding to each new noisy realization of the original signal.

It must be noticed that these preliminary EEMD results could only be compared with those shown in Fig. 5. Usually, in the EEMD approach the number of realizations is high ($200 \leq N_e \leq 1000$) while the sifting number is fixed as low as
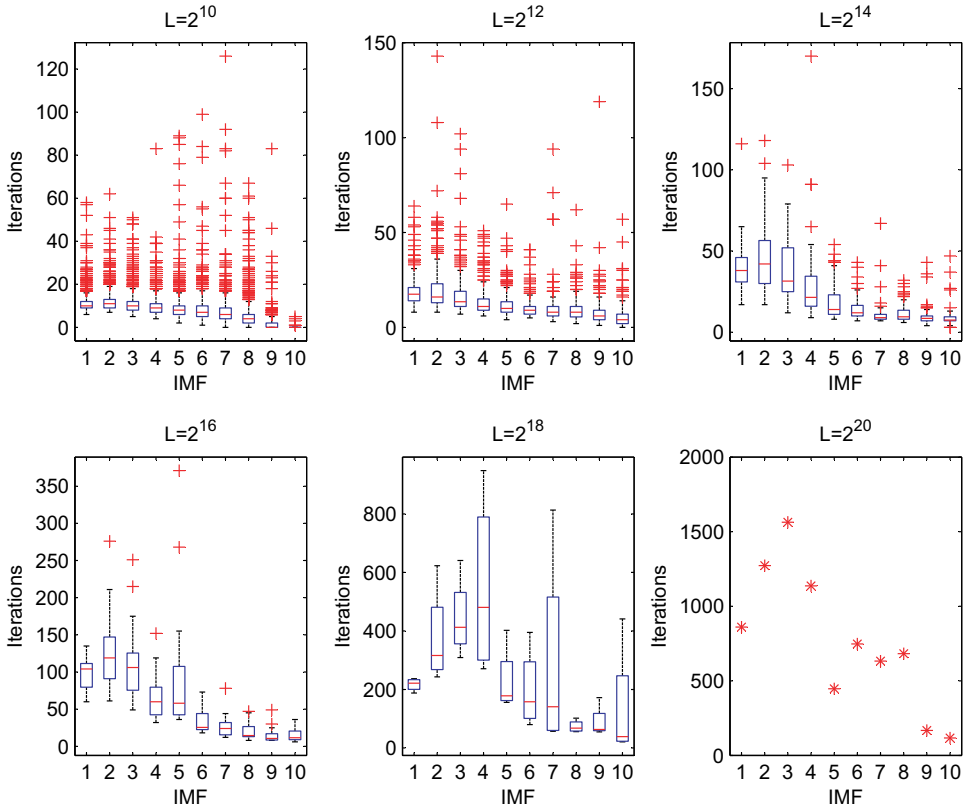
Fig. 7.   Boxplots of the number of sifting iterations for data of lengths $L = 2^{10}, 2^{12}, \ldots, 2^{10}$ and number of sifting iterations for $L = 2^{20}$ datapoints. The maximum number of iterations was unlimited.

possible. Therefore, taking into account its extensive computational cost, it would not be a realistic situation to perform experiments in the EEMD framework with unlimited number of iterations. Therefore, the new experiments with EEMD should be carefully designed. These ideas will be explored and developed in future works.

4. Conclusions

A study of the statistical properties of the IMFs obtained by EMD of Gaussian noise, and its dependence on data lengths and number of allowed sifting iterations was presented. It indicates that the PDFs is strongly relied on the data length, and on the maximum allowed number of iterations. Depending on these parameters, the PDF can have a Laplacian alike or a multimodal shape. Only in a few settings, a Gaussian distribution is obtained. These results were confirmed by two different normality tests. Similar conclusions were derived from uniformly distributed white noise data.

Acknowledgments

References

1. N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung and H. H. Liu, The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. Roy. Soc. London A* **454** (1998) 903–995.
2. P. Flandrin, G. Rilling and P. Gonçalvès, Empirical mode decomposition as a filter bank, *IEEE Signal Process. Lett.* **11** (2004) 112–114.
3. P. Flandrin and P. Gonçalvès, Empirical mode decomposition as data-driven wavelet-like expansions, *Int. J. Wavelets Multiresolution Inform. Process.* **2** (2004) 477–496.
4. Z. Wu and N. E. Huang, A study of the characteristics of white noise using the empirical mode decomposition method, *Proc. Roy. Soc. London A* **460** (2004) 1597–1611.
5. Z. Wu and N. E. Huang, Statistical significance test of intrinsic mode functions, in *Hilbert–Huang Transform and its Applications*, in *Interdisciplinary Mathematical Science*, Vol. 5 (World Scientific, Singapore, 2005), pp. 107–127.
6. C. M. Jarque and A. L. Bera, A test for normality of observations and regression residuals, *Int. Stat. Rev.* **55** (1987) 163–172.
7. G. Rilling, P. Flandrin and P. Gonçalvès, On Empirical Mode Decomposition and its algorithms, in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, NSIP-03*, Grado (I), June 2003.
8. http://perso.ens-lyon.fr/patrick.flandrin/emd.html
9. A. W. Bowman and A. Azzalini, Applied smoothing techniques for data analysis, in *The Kernel Approach with S-plus Illustrations* (Oxford University Press, New York, 1997).
10. Z. Wu and N. E. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method, *Adv. Adapt. Data Anal.* **1**(1) (2009) 1–41.