MULTISCALE ZERO-CROSSING STATISTICS OF INTRINSIC MODE FUNCTIONS FOR WHITE GAUSSIAN NOISE

Süleyman Baykut , Tayfun Akgül

Department of Electronics and Communications Engineering, Istanbul Technical University,
34469, Istanbul, Turkey
phone: + (90) 212 285 3568, email: baykut@itu.edu.tr, tayfunakgul@itu.edu.tr

ABSTRACT

In this paper, the statistical characteristics of zero-crossing intervals and zero-crossing amplitudes (the time interval and the absolute value of extrema between two successive zero-crossings) of intrinsic mode functions of white Gaussian noise are studied. Intrinsic mode functions are extracted by empirical mode decomposition method. Numerous simulations are conducted and the probability distribution functions of zero-crossing intervals and amplitudes are obtained. Simulation results are included. These findings are important to determine the statistical significance of IMFs. These white noise-only case statistical characteristics can be used for signal detection and/or signal/noise separation and an efficient noise reduction can be achieved.

1. INTRODUCTION

Zero-crossings (ZCs) reveal rich information about the signal characteristics with relatively low computational load [1, 2]. For example, ZC counts in a signal gives coarse but fast estimation of the fundamental frequency without requiring any spectral analysis. ZCs are among the most meaningful features in a signal. Statistics of ZCs of random functions, such as normally distributed, band-limited noise, have been studied starting from early 40's [3]. Due to a lack of analytical expressions of the ZC statistics, this is still an ongoing research topic in diverse fields [4, 5].

It is relatively easier and meaningful to characterize the signal by the ZC statistics when the signal has narrow and band-pass spectrum, which is the case in Intrinsic Mode Functions (IMF) of white noise. IMFs are obtained by Empirical Mode Decomposition (EMD) which is a data-driven and iterative method [6]. In this study, instead of the noise signal itself, we investigate the statistical characteristics of zero-crossings of IMFs of white noise via simulations. The characterization of the ZCs of IMFs can be used in various applications such as noise reduction or signal detection in noise. It can also be used to determine the IMF significance level which is important to assign the IMFs corresponds to signal itself or to noise component.

Statistics of IMF amplitudes of white noise and fractional Gaussian noise are studied in [7] and in [8, 9]. Since EMD is an empirical method with no compact analytical definition, most of the studies are done by numerical simulations. The results are important for better understanding of the EMD procedure. These results can also be used for denoising or detrending of noisy signals [10] by comparing the IMF energies along scales with the ones obtained from the noise-only case.

S. Baykut's work is partly supported by TÜBİTAK-BİDEB.

In this paper, instead of treating the intrinsic mode as a whole (the mode energies), first, these modes are divided into small intervals, namely ZC intervals, and the statistical behavior of these intervals is investigated. Two ZC vectors are formed; 1) Zero Crossing Interval (ZCI) (the time difference between two successive zero-crossings), 2) Zero Crossing Amplitude (ZCA) (the absolute value of the extrema point between two successive zero-crossings). Simulations are carried out to characterize the statistical behaviors (the distributions and the relationship between ZCI and ZCA) of these vectors for normally distributed white noise. Once the characteristics are determined, they can be used for a noisy signal to label each ZC intervals as noise-component or signal-component which then yield a convenient noise/signal separation.

In Section 2, EMD algorithm is briefly explained and ZCs are defined. Experiments and observations are given in Section 3 and the paper is concluded in Section 4.

2. BACKGROUND

2.1 Empirical Mode Decomposition

EMD is a signal decomposition methods that extracts the intrinsic components from the signal without using any a priori fixed basis as in Fourier or wavelet analysis [6]. It is an ideal tool to analyze nonstationary and nonlinear processes. The main procedure of EMD is called "sifting procedure", where the iterations isolate the fast oscillations locally in time. After the first sifting procedure, the first IMF is obtained as a waveform which contains the signal details at the finest time scales (highest frequencies). This IMF is subtracted from the original signal, and the steps are repeated on the remainder signal in order to obtain the next IMF. The same procedure is applied iteratively until some stopping criteria are satisfied [11]. IMFs are narrow-band, zero-mean signals where the number of extrema points and the number of zero-crossings are equal or differ by 1 at most. The signal $x(t)$ is decomposed into K IMFs by EMD as:

$$x(t) = \sum_{i=1}^{K} d_k(t) + r(t) \tag{1}$$

Here, $d_k(t)$ is the k^{th} IMF and $r(t)$ is the residue signal. Each IMF lays at lower frequency regions locally in time-frequency domain than the previous one.

EMD acts as a dyadic filter bank for white noise [7, 9]. It separates the white noise into intrinsic modes (IMFs) where the mean periods are increasing by the power of 2 as the scale decreases.

 ISSN 2076-1465

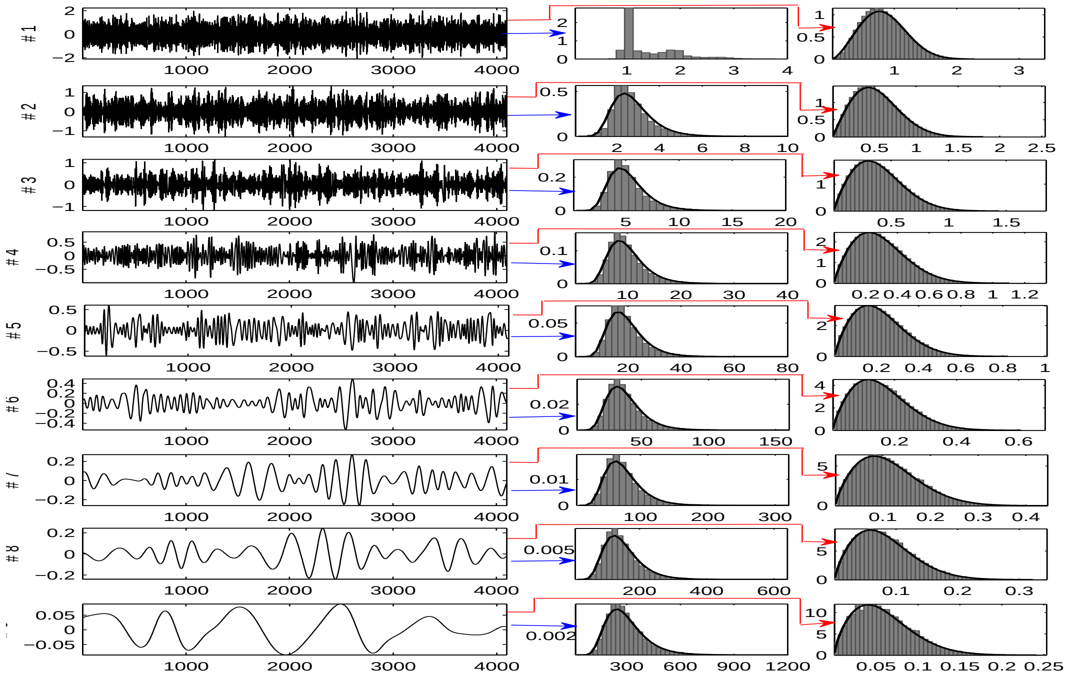

Figure 1: 1^{st} Column:IMFs of single realization of white noise; 2^{nd} Column: ZCI histograms and the lognormal fits; 3^{rd} Column: ZCA histograms and the Weibull fits (x-axis is in terms of number of data points).

2.2 Zero-crossings

It is already shown that IMFs of white noise are normally distributed [7, 9]. In our work, apart from the IMF amplitude distribution, we experimentally examine the statistics of the zero-crossing intervals and the zero-crossing amplitudes of IMFs of white noise.

Let $z_k[i]$ be the i^{th} ZC instant of k^{th} IMF ($d_k(t)$):

$$z_k[i] = \{t_i | d_k(t_i) = 0\}, \qquad i = 1, 2, ..., \mathscr{Z}_k \tag{2}$$

where $\mathscr{Z}_k$ is the total number of ZCs at k^{th} IMF. Then, ZC intervals $\tau_k[i]$ are expressed as:

$$\tau_k[i] = z_k[i+1] - z_k[i] \tag{3}$$

where the ZC amplitudes $a_k[i]$ are expressed as below:

$$a_k[i] = max(|d_k(t)|) \;\mid\; t \in [z_k[i], z_k[i+1]] \tag{4}$$

for $i = 1, 2, ..., \mathscr{Z}_k - 1$.

As stated in [12], the positions of multi-scale zero-crossings may provide information of the signal characteristics. However, such representation is not stable unless a measure regarding the size of the signal between successive zero crossings are not provided. Therefore, in this paper, besides the ZC intervals, ZC amplitudes are also taken into consideration as a size measure.

3. EXPERIMENTS AND OBSERVATIONS

In order to establish the statistical characteristics of ZCIs and ZCAs of IMFs of white noise, numerical simulations are conducted. For the experiments, $M = 1000$ independent white noise processes are synthesized with the length of $N = 2^{12}$. The IMFs of each realizations are obtained by EMD algorithm referred as $d_k^{(m)}[n]$ (with $m = 1, ...M$; $n = 1, ...N$ and $k = 1, ...K_m$). The number of IMFs, K_m, in a wide band signal equals to $log_2 N$ approximately [11]. However, for the selected data length, the algorithm yields different number of IMFs for each realization of noise varying between 9-12. Higher index IMFs covers very narrow band at low frequency region and have negligible amplitudes compared to the ones with lower indices. Therefore, only the first 9 IMFs are taken into account for the analysis for each realization. The first 9 IMFs of a single realization of white noise are plotted in the 1^{st} column of Fig. 1. The zero-crossing points[1] are determined by finding the sign-change-points in the signals, then ZCI and ZCA signals are constructed for each IMF.

3.1 ZCI Distributions

ZCIs of each specific IMF obtained from different realizations are grouped in order to construct the main ZCI vector. The histograms of ZCIs are obtained and then individually normalized so that the area under the histogram bars equals to unity. In Fig. 1, the normalized histograms of ZCIs are plotted in the 2^{nd} column. Several distribution functions (i.e., Normal, Lognormal, Rayleigh, Weibull, Gamma, Beta, Extreme Value, T Location-Scale, etc.) are fit to histograms and it is observed that lognormal distribution gives the minimum fit error, i.e., each ZCI data (except the ones that belong to the 1^{st} IMF, explained in the next paragraph) approximately fits the lognormal distribution. The fitted lognormal density functions can be seen in the same plots as solid lines. IMFs at lower frequency regions (higher indices) have less number of zero crossings, so the distribution of these ZCIs are less

[1] Note that, since we deal with discrete time signals, the signal does not always necessarily equals zero at a zero-crossing point. The sign of the signal can change from minus(plus) to plus(minus) without having a value of zero. So the zero-crossing points can be determined by using a simple linear fit between two points where the signal changes sign.

smooth which cause a deviation from lognormal distribution. If the data length is long enough, the ZCI sequences of higher indice IMFs will behave similar to lower indice IMFs.

The normalized histogram of ZCIs of the 1^{st} IMF has 2 peaks as shown in the first plot of 2^{nd} column in Fig. 1. ZCIs of the 1^{st} IMF behave different than the ones belong to other IMFs. This difference is basically coming from different power spectrum behavior of the 1^{st} IMF. It is known that IMFs' power spectra have band-pass shape which are identical on semi-logarithmic scale [8], but the spectrum of 1^{st} IMF has a high-pass shape that covers higher half of the frequency region. Furthermore, there is non-negligible leakage to the lower frequency bands that causes a wide range of ZCIs.

Mean periods of IMFs increase by power 2 as the IMF order increases [8, 9]. This relationship evokes a scaling property between ZCI density functions ($p_k(\tau)$) of IMFs:

$$p_{\acute{k}}(\tau) = \rho^{(k-\acute{k})} p_k(\rho^{(k-\acute{k})}\tau) \tag{5}$$

Here the scaling factor ρ is approximately 2. In Fig. 2, normalized pdf of ZCIs for $9 \geq \acute{k} > k \geq 2$ are given where they approximately overlap. The shape parameters of lognormal fits are the same and the scale parameters differ by $ln(2)$.

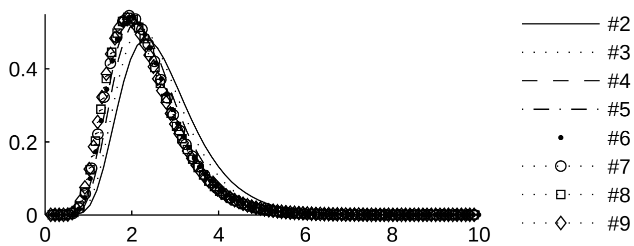

Figure 2: Normalized ZCI lognormal pdfs (x-axis is in terms of number of data points).

3.2 ZCA Distributions

In this section the distribution functions of ZCAs are studied. Determination of the statistical characteristics of maximum (or minimum) points of some real data has critical importance such as sea surface wave heights used in coastal engineering. For example in [13], the distribution of the maxima of the sea-surface displacement is derived which approximates to Rayleigh distribution when the signal is normally distributed and has a narrow-band spectrum. Since IMFs obtained from white noise are also normally distributed and narrow-band, a Rayleigh-like distribution is expected.

ZCA vectors are constructed in a similar fashion as ZCIs defined above. The normalized histograms of ZCAs are given in the 3^{rd} column in Fig. 1. In these data sets, Weibull distribution has the minimum fit. Although the fit error is higher in the 1^{st}IMF compared to the remaining IMFs, the distribution still obeys the Weibull distribution.

3.3 ZCI vs ZCA

In this section, the relationship between ZCI and ZCA for each IMF is investigated. In Fig. 3, logarithm of ZCAs are plotted versus logarithm of ZCIs. Point groups in the same color are the scattered distributions of each ZCI and ZCA of a particular IMF. The mean of each group is also shown in Fig. 3 as circles in order to observe the relationship. Through visual inspection one can suggest that there is approximately linear relationships along scales:

$$log_2(\overline{a_k}) \simeq \alpha log_2(\overline{\tau_k}) + c \tag{6}$$

where c is a constant. When a line is fitted to the data, the slope of the fitted line is estimated as -0.48214 (which is close to -0.5). It is stated in [7] that there is a simple linear relation between IMF energies, E_k and average IMF period, T_k as:

$$log_2 E_k = b - log_2 T_k \tag{7}$$

where b is constant. Since IMFs are sinusoid-like signals, we can approximate the energy of each IMF as "$\frac{\overline{a_k}^2}{2}$" and the average IMF period as "$\overline{\tau_k}$" where $\overline{a_k}$ and $\overline{\tau_k}$ are the means of ZCA and ZCI respectively. These approximations in equation (7) yields:

$$log_2(\overline{a_k}) = \frac{b-1}{2} - \frac{1}{2} log_2(\overline{\tau_k}) \tag{8}$$

Note that we obtain the relationship in equation (6) with $\alpha = -0.5$ and $c = \frac{b-1}{2}$. The slope of the straight line becomes -0.5 (which is already found as -0.48 by linear fit in Fig. 3).

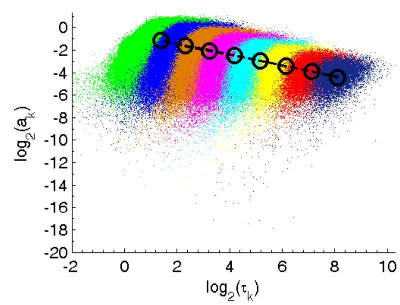

Figure 3: ZCI versus ZCA in logarithmic scale (The mean values of each group as black circles. The slope of the linear fit is estimated as -0.48214).

3.4 Periodic Signal plus Noise Example

When the signal of interest consists of both noise and some periodic components, the ZCI distributions reveal anomaly regarding "corruption" at specific IMFs. This is shown in a simple example where the signal is white noise corrupted a sinusoidal signal with the angular frequency $\omega_0 = 0.04\pi$ where the signal-to-noise ratio is 5dB. Obtained IMFs are plotted in the 1^{st} column in Fig. 4. The ZCI and ZCA distributions of 2^{nd} to 5^{th} IMFs are given in Fig. 4 in the 2^{nd} and the 3^{rd} column, respectively. The lognormal and Weibull fits for noise-only case are also shown in the figures as black lines for comparison.

Deviations from lognormal distribution for ZCI and deviations from Weibull distribution for ZCA are observed at 4^{th} and 5^{th} IMFs, which means that the periodical signal components are present in these two IMFs. Here, the periodical signal is broken into 2 IMFs because of "mode mixing" phenomenon which means that different modes of oscillations coexist in a single IMF. Clearly, by using these histogram, one can create a simple algorithm to determine which ZC interval belongs to signal or to noise by using these histograms.

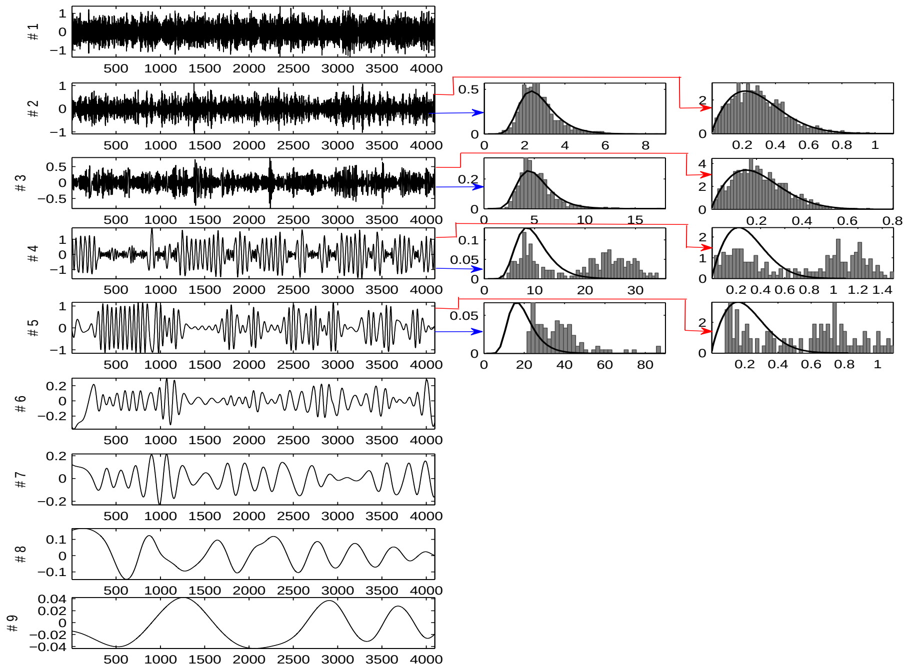

Figure 4: 1^{st} Column:IMFs of White Noise + $cos(0.04\pi t)$; 2^{nd} Column: ZCI histograms of 2^{nd} to 5^{th} IMFs (the lognormal fits of noise-only case is also shown as black lines for comparison); 3^{rd} Column: ZCA histograms of 2^{nd} to 5^{th} IMFs (the Weibull fits of noise-only case is also shown as black lines for comparison) (x-axis is in terms of number of data points).

4. CONCLUSION

In this paper, the distribution of ZCIs and ZCAs of white Gaussian noise IMFs are obtained by numerical analyses. It is observed that ZCIs (except IMF#1) obey lognormal distribution whereas ZCAs have Weibull distribution. When the data contains signal with additive white Gaussian noise, deviations from lognormal and Weibull distributions are observed in some IMFs which can be used to detect and separate the signal and noise components. It can also be used to determine the IMF significance level (as a measure) which is important to assign the IMFs correspond to signal itself or to noise component. Further explorations are needed for other types of noise.

REFERENCES

[1] B. Kedem, "Spectral analysis and discriminant by zero-crossings," *Proc. of the IEEE*, vol. 74, no. 11, 1986.

[2] T. V. Sreenivas and R. J. Niederjohn, "Zero-crossing based spectral analysis and svd spectral analysis for formant frequency estimation in noise," *IEEE Transactions on Signal Processing*, vol. 40, no. 2, Feb. 1992.

[3] O. S. Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 23, pp. 282–332, 1944.

[4] F. Abdellani, G. Rasigni, M. Rasigni, and A. Lebaria, "Distribution of zero crossings for the profile of random rough surfaces," *Applied Optics*, vol. 31, no. 22, pp. 4534–4539, 1992.

[5] K. I. Hopcraft, P. C. Ingrey, and E. Jakeman, "Power-law-distributed level crossings define fractal behavior," *Physical Review E*, vol. 76, no. 031134, pp. 1–11, 2007.

[6] 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 nonstationary time series analysis," *Proc. Roy. Soc. London A*, vol. 454, pp. 903–995, 1998.

[7] 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*, vol. 460, pp. 1597–1611, 2004.

[8] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.

[9] P. Flandrin and P. Goncalves, "Empirical mode decomposition as data-driven wavelet-like expansions," *Int. Jrn. of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 4, pp. 1–20, 2004.

[10] P. Flandrin, P. Goncalves, and G. Rilling, "Detrending and denoising with empirical mode decompositions," in *Proceeding of EUSIPCO-04*, Wien, Austria, September 2004, pp. 1581–1584.

[11] G. Rilling, P. Flandrin, and P. Goncalves, "On empirical mode decomposition and its algorithms," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*, Grado, 2003.

[12] S. Mallat, "Zero-crossings of a wavelet transform," *IEEE Transactions on Information Theory*, vol. 37, no. 4, pp. 1019–1033, 1991.

[13] D. E. Cartwright and M. S. Longuet-Higgins, "The statistical distribution of the maxima of a random function," *Proc. Roy. Soc. London*, vol. 237, no. 1209, pp. 212–232, 1956.