Theoretical Basics of the EMD: Example of Two Harmonics

Michael Feldman[1]
Technion – Israel Institute of Technology,
http://hitech.technion.ac.il/feldman/EMD_theory.pdf

1 Introduction

The Empirical Mode Decomposition (EMD) technique, an original technique first introduced by Huang et al. [1], adaptively decomposes a signal into the simplest intrinsic oscillatory modes (components). The method is able to visualize signal energy spread between available frequencies locally in time, thus resembling the wavelet transform. Hence, it was immediately applied in diverse areas of signal processing. In the field of mechanical sound systems and vibration, the EMD method has also been applied widely for diagnostics and structural health monitoring, as well as in analysis and identification of nonlinear vibration, mainly in rotating systems with typical elements such as bearings and gears.

In parallel, sophisticated studies devoted to analyzing the essential shortcomings of the EMD and its restrictions in comparison with other decomposition methods began to appear. One of the first limitations found was the method's rather low frequency resolution [2] meaning that the EMD can resolve only distant spectral components differing by more than octave. Another weak point of the method was that it receives false artificial components not present in the initial composition. However, the newest Ensemble Empirical Mode Decomposition (EEMD) method [3] largely overcomes the false mode mixing problem of the original EMD and provides physically unique decompositions.

All earlier publications agree that the EMD is defined only empirically by its algorithm, and does not for the moment set out an analytical formulation that would allow for theoretical analysis and analytical performance evaluation [4] . An exception is the typical case of a composition of a harmonics and a slow varying aperiodic (like DC) trend. The Hilbert Transform projection of such a composition looks even simpler than the initial signal because the HT of the constant is equal to zero. Therefore, it is simple to show analytically how the EMD removed the slow trend from the composition.

These questions are considered in depth in the work by G. Rilling and P. Flandrin [5] , which considers the case of decomposing two harmonics. The work provides theoretical and experimental proofs for the existence of three domains of amplitude-frequency harmonics relations: 1) the components are well separated and correctly identified; 2) the harmonics are considered as a single waveform, and 3) the EMD

[1] Cite: [8]; MFeldman@technion.ac.il

does something else. Nevertheless, this work is theoretically constructed based upon rather complicated Fourier transform models [5] .

The notion of such two-tone signal by itself appears to be rather fruitful. A most realistic vibration or sound signal can be thought of as consisting of a linear combination of two or more sinusoids. In the present study we consider the theoretical foundation of the EMD decomposition in the simplest way through direct analysis of the harmonics separation. For the sum of two harmonics, we demonstrate why the low frequency harmonics remains while the high frequency harmonics is sifted out first in the EMD procedure. We also show a theoretical limiting frequency resolution achieved by the method.

2 Analytic signal

2.1 Signal envelope

The interpretation of an envelope, defined as the absolute value (modulus or magnitude) of a complex (typically analytic) representation of a signal, has been the subject of investigation for some time. For general modulated signals, it is often convenient to define the analytic signal $X(t) = x(t) + i\tilde{x}(t),$ where $\tilde{x}(t)$ is related to $x(t)$ by the Hilbert Transform (HT). According to analytic signal theory, a real vibration process $x(t)$ measured by a transducer, for example, is only one of the possible projections (the real part) of some analytic signal $X(t)$. Then, the second or quadrature projection of the same signal (the imaginary part $\tilde{x}(t)$) will be conjugated according to the HT. The analytic signal is represented geometrically in the form of a phasor rotating in a complex plane. The traditional representation of the analytic signal in its trigonometric or exponential form,

$$X(t) = x(t) + i\tilde{x}(t) = |X(t)|\left[\cos\varphi(t) + i\sin\varphi(t)\right] = A(t)e^{i\varphi(t)}, \tag{1}$$

can be used to determine its instantaneous amplitude (envelope, magnitude, absolute value)

$$A(t) = |X(t)| = \sqrt{x^2(t) + \tilde{x}^2(t)} = e^{\mathrm{Re}[\ln X(t)]} \tag{2}$$

and its instantaneous phase

$$\psi(t) = \arctan\frac{\tilde{x}(t)}{x(t)} = \mathrm{Im}\left[\ln X(t)\right], \tag{3}$$

where $\tilde{x}(t)$ is the Hilbert transform of $x(t)$ that can be written as the convolution integral of $x(t)$ with $1/\pi t$ as $\tilde{x}(t) = x(t) * 1/\pi t$. The plus sign of the root square (Eq.(2)) corresponds to the **upper envelope** and the minus sign corresponds to the opposite sign **lower envelope**, so they are always in antiphase relation.

2.2 Instantaneous frequency

The first derivative of the instantaneous phase (Eq.(3)) as a function of time $\omega(t) = \dot{\psi}(t)$, called the instantaneous angular frequency, plays an important role. The angular frequency $\omega(t) = 2\pi f(t)$ has the

radian-per-second dimension and the cycle frequency $f(t)$ has the Hertz dimension. The whole phase-unwrapping problem can be avoided simply when finding the instantaneous frequency (IF) by differentiation of the signal itself

$$\omega(t) = \frac{x(t)\dot{\tilde{x}}(t) - \dot{x}(t)\tilde{x}(t)}{A^2(t)} = \mathrm{Im}\left[\frac{\dot{X}(t)}{X(t)}\right] \tag{4}$$

2.3 First derivative of the signal

The first derivative (differentiating the function once) yields the slope of the tangent to that function, and the analytic signal notion (Eq.(1)) allows us to describe a relationship between an initial complex signal and its first derivative as

$$\dot{X} = X\left(\frac{\dot{A}}{A} + i\omega\right). \tag{5}$$

Every **maximum (top extremum)** and **minimum (bottom extremum)** point of the function, known collectively as a set of the local extrema points x_{extr}, is uniquely defined by the first derivative when the slope of the tangent is equal to zero. The initial signal and its envelope have common tangents at points of contact, but the signal never crosses the envelope. The common points of the contact between the signal and its envelope do not always correspond to the local extrema of a multi-component signal. The local extrema always have a zero tangent slope, but the common points of the contact can have nonzero value of the tangent slope. The distance between the common points of contact and the extrema points of the signal plays a dominant role in explaining the EMD mechanism. In effect, without such a difference between the envelope and the local extrema, the sum of maxima and minima curves desired by the EMD would be always equal to zero, just like the zero sum of the upper and the lower envelopes.

3 Distance between envelope and extrema

3.1 Points of contact between envelope and signal

Both the envelope $A(t)$ and the initial function $x(t)$ can be functions that vary over time. Their relation has a known and simple form: $x(t) = A(t)\cos\left[\varphi(t)\right]$. During variation, the initial function and its envelope will have common contact points in time where these functions touch each other. The function and the envelope have common tangents at these points of contact. The condition of their contact takes the form:

$$x(t) = \pm A(t),\ \cos\left[\varphi(t)\right] = \pm 1,\ \varphi(t) = \begin{cases} 0 \\ \pm\pi \end{cases}, \tag{6}$$

indicating that the common contact points are always located on the envelope.

3.2 Local extrema points

Unlike the common contact points, the local extrema depend on the zero slope of the first derivative. The signal's derivative can be expressed as

$$\dot{X}(t) = \sqrt{\dot{A}^2(t) + A^2(t)\omega(t)^2}\exp\left\{i\left[\varphi(t) + \arctan\frac{A(t)\omega(t)}{\dot{A}(t)}\right]\right\}. \tag{7}$$

The derivative (Eq.(7)) introduces a new varying velocity envelope and a new varying velocity phase function. Let us find points in time at which the slope of the tangent is zero and at which the function reaches its extrema. These points correspond to the zero value of the cosine function of the new varying velocity phase angle (Eq. (7)): $\varphi(t) + \arctan\frac{A(t)\omega(t)}{\dot{A}(t)} = \pm\frac{\pi}{2}$,

$$\varphi(t) = \pm\frac{\pi}{2} - \arctan\frac{A(t)\omega(t)}{\dot{A}(t)}. \tag{8}$$

Notice that in the case of a monoharmonic signal, the envelope is constant ($\dot{A}(t) = 0$) and the conditions in Eq.(6) and Eq.(8) become identical: $\varphi(t) = 0$, or $\varphi(t) = \pm\pi$. For the monoharmonics, the local extrema always lie on the envelope.

3.3 Deviation of local extrema from envelope

In the general case, the vertical position of the local extrema $x_{extr}(t) = A(t)\cos\left[\varphi(t)\right]$ are determined by the cosine projection of the new velocity phase (Eq.(8)):

$$x_{extr}(t) = A(t)\cos\left[\varphi(t)\right] = \pm\frac{A^2(t)\omega(t)}{\dot{A}(t)\sqrt{1 + \frac{A^2(t)\omega^2(t)}{\dot{A}^2(t)}}} \tag{9}$$

The obtained continuous vertical position of the local extrema differs from the envelope function. Generally, the cosine projection can assume values from 1 through 0 up to -1. Therefore the corresponding local maxima can be equal to the envelope value, be as small as zero, or even take negative values. The cosine projection is controlled by the variable $\frac{A(t)\omega(t)}{\dot{A}(t)}$ whose shape, level, and frequencies depend on the initial signal $x(t)$. In turn, the variable $\frac{A(t)\omega(t)}{\dot{A}(t)}$ is determined by the relation between the nominator and the denominator.

For small envelope variations when $\dot{A}_{\max} \ll (A\omega)_{\max}$, the cosine projection $\cos\left[\varphi(t)\right]$ practically does not differ from 1. This condition always forces the local maxima to be on the envelope itself. That is, the connected local maxima and the connected local minima curves just repeat the corresponding and opposite signed upper and lower envelopes. For larger envelope variations, the cosine projection during oscillation

can decrease to zero or even to negative values. It will produce zero and negative local maxima below zero up to the opposite signed lower envelope.

3.4 Local extrema sampling

Connected together, the local maxima give shape to the maxima curve, and the correspondingly connected local minima give shape to the minima curve. As shown in Eq.(9), the vertical values of both the top and the bottom extrema curves are generated by the continuous function of multiplying the envelope by the cosine projection. However, the discrete extrema points themselves are formed by digitizing the continuous cosine projection (Eq.(9)) at distinct moments of time. These sampling moments completely depend on the instantaneous frequency $\omega(t)$ of the initial signal. Thus, from the continuous function of the cosine projection $x_{extr}(t)$ (Eq.(9)) we have a set of the extrema sampled with the instantaneous frequency $\omega(t)$. The series of the sampled extrema interpolated by spline generate two continuous profiles (extrema curves) required by the EMD method [1] .

The obtained discrete set of samples does not repeat the original continuous function $x_{extr}(t)$. If the frequency of the continuous function (Eq.(9)) exceeds (overlaps) the Nyquist frequency $0.5\omega(t)$, the sampled extrema line undergoes aliasing with a new folding frequency of around half the sampling frequency: $\omega_{fold}(t) = \left|\omega(t) - \omega_{x_{extr}}(t)\right|$. But if the frequency of the continuous function (Eq.(9)) lies below the Nyquist frequency $0.5\omega(t)$, no aliasing occurs, and the sampled extrema curves will follow the the signal envelope.

4 Decomposition of two harmonics

4.1 Envelope of two harmonics

A case of combination of two harmonics is the rather obvious and interesting case of signal composition [5] . This case enables us to discover and prove some important features of the EMD. If a signal composition is a sum of two harmonics: $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$, the envelope $A(t)$ of the double-component signal composition can be written as Eq.(2):

$$A(t) = \left[A_1^2 + A_2^2 + 2A_1A_2 \cos\left(\omega_2 - \omega_1\right)t \right]^{1/2}. \tag{10}$$

The signal envelope $A(t)$ consists of two different parts, that is, a slowly varying part including the sum of the component amplitudes squared and a rapidly varying part, oscillating with a new frequency equal to the difference between the component frequencies.

4.2 Instantaneous frequency of two harmonics

The IF $\omega(t)$ of the double-component composition according to Eq. (4) (for definiteness $A_1 > A_2$; $\omega_2 > \omega_1$) is:

$$\omega(t) = \omega_1 + \frac{(\omega_2 - \omega_1)\left[A_2^2 + A_1 A_2 \cos\left(\omega_2 - \omega_1\right)t\right]}{A^2(t)} \tag{11}$$

The IF of the two tones considered in Eq.(11) is generally time-varying and exhibits asymmetrical deviations about the frequency ω_1 of the largest harmonics. Not only does the IF for two tones have time-varying deviations, but these deviations always force the IF beyond the frequency range of the signal components. The IF in principle consists of two different parts, that is, a frequency of the first largest component ω_1 and a rapidly varying asymmetrical oscillating part. For large amplitude of the second harmonics when

$$\frac{A_2}{A_1} > \frac{\omega_1}{\omega_2}, \quad \text{or} \quad A_2 > \omega_2^{-1} \text{ when } A_1 = 1, \omega_1 = 1, \tag{12}$$

the IF of the composition becomes negative.

Appearance of a negative IF corresponds to the arrival of the local negative maximum or local positive minimum of the signal. The upper tangent to the negative maximum touches the signal lower envelope, and, vice versa, the lower tangent to the positive minimum touches the signal upper envelope. The appearance of these extrema with opposite signs increases the local extrema deviation from the envelope.

4.3 Average instantaneous frequency

Thus, Eq.(11) shows that the IF consists of two different parts, that is, a slow varying frequency of the first component ω_1 and a rapidly varying asymmetrical oscillating part. However, the rapidly varying asymmetrical oscillating part of the IF has an important feature. If we now integrate the oscillating part with the integration limits corresponding to the full period of the difference frequency $\left[0 \quad T = \frac{2\pi}{\omega_2 - \omega_1}\right]$,

$$\int_0^T \frac{(\omega_2 - \omega_1)\left[a_2^2 + a_1 a_2 \cos\left(\int(\omega_2 - \omega_1)dt\right)\right]}{a^2(t)} dt = 0, \tag{13}$$

we get the definite integral equal to zero [6] . This means that the average value or the first moment of the IF (Eq.(11)) is just equal to the frequency of the largest harmonics $\langle\omega(t)\rangle = \omega_1(t) + \int_0^T \omega(t) = \omega_1(t) + 0$. This important property of the IF offers the simplest and most direct way of estimating the mean frequency of an a priori unknown composition with the largest signal component.

4.4 Distance between envelope and extrema

The vertical position of the local extrema according to Eq.(9) depends on the variable $\frac{A(t)\omega(t)}{\dot{A}(t)}$, which in the case of two harmonics has the same period as the period of the envelope variation $2\pi/(\omega_2-\omega_1)$: $\frac{A(t)\omega(t)}{\dot{A}(t)} = -\frac{\omega_1 A_1/A_2 + \omega_2 A_2/A_1}{(\omega_2-\omega_1)\sin(\omega_2 t-\omega_1 t)} - \frac{\omega_2+\omega_1}{\omega_2-\omega_1}\cot(\omega_2 t-\omega_1 t)$. Multiplying the envelope and the cosine projection (Eq. (9)) generates the vertical position as a periodic function with the same period $2\pi/(\omega_2-\omega_1)$:

$$x_{extr}(t) = A(t)\frac{A(t)\omega(t)}{\dot{A}(t)\sqrt{1+\frac{A^2(t)\omega^2(t)}{\dot{A}^2(t)}}}. \qquad (14)$$

This vertical position of the local maxima during a single period varies from its highest to the lowest position, thus specifying a band with all possible local extrema.

When the instantaneous frequency of the composition becomes negative $\frac{A_2}{A_1} > \frac{\omega_1}{\omega_2}$, the vertical position is a monotonic function with a top maximum always equal to the sum of amplitudes of both harmonics: $x_{\max(top)}(0) = A_1 + A_2$ and with a bottom maximum position equal to the difference of amplitude that has the negative value (Figure 1, a): $x_{\max(bottom)} = -A_1 + A_2$.

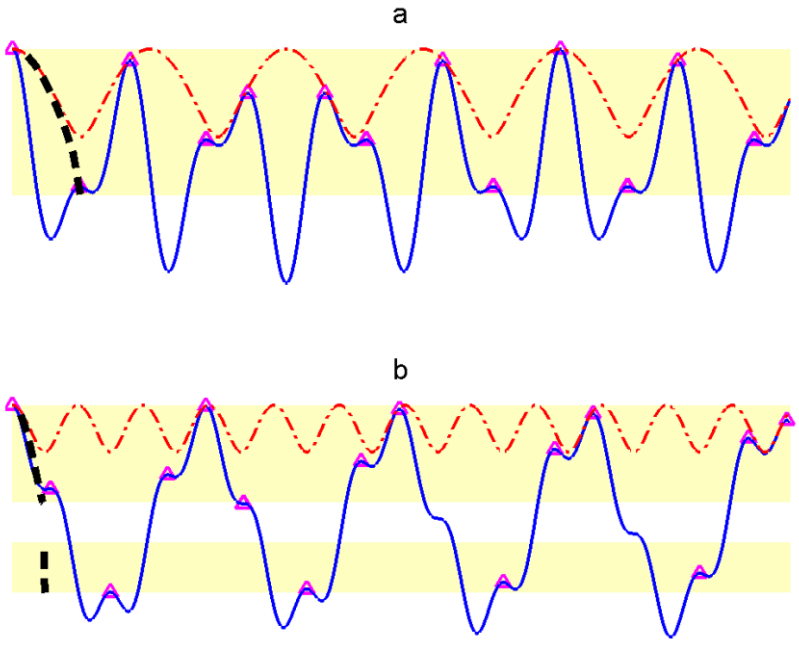

Figure 1.

The vertical position of the local maxima (**--**), the initial signal (—), the upper envelope (--) and the top maxima (Δ): (a) the negative instantaneous frequency ($A_1=1, \omega_1=1, A_2=0.6, \omega_2=1.8$), (b) the positive instantaneous frequency ($A_1=1, \omega_1=1, A_2=0.25, \omega_2=3.9$).

For an instantaneous frequency that is always positive, when $\frac{A_2}{A_1} < \frac{\omega_1}{\omega_2}$, the top vertical position of maxima is again equal to the sum of amplitudes: $x_{\max(\text{top})}(0) = A_1 + A_2$. But now the vertical position decreases monotonically only until the intermediate bottom position $x_{\max(\text{bottom}(1))} = \left(A_1^2 + A_2^2 - A_1^2\,\omega_1^2/\omega_2^2 - A_2^2\,\omega_2^2/\omega_1^2\right)^{\frac{1}{2}}$ (Figure 1, b). During the remainder of the period, the resultant vertical position jumps down to the negative symmetric value $-x_{\max(\text{bottom}(1))}$, specifying a second isolated band with the negative local maxima. The vertical position for the second band at the end of the period monotonically continues to the negative extreme bottom position $x_{\max(\text{bottom})} = -A_1 + A_2$ (Figure 1, b). Theoretically, the values of the extrema deviation from the envelope depend on two ratios: the envelope, and the frequency of the harmonics. For a very small amplitude of the second harmonics $A_2 \le 0.3A_1\omega_1/\omega_2$, this intermediate bottom position practically does not differ from the smallest envelope value $A_1 - A_2$. This means that the positive maxima points of the first band always will lie on the envelope. For the other ratio of the harmonics parameters $0.3A_1\omega_1/\omega_2 \le A_2 < A_1\omega_1/\omega_2$, the local maxima will wander vertically more and more from the envelope.

4.5 Mean value between the local maxima and minima curves

In general, the initial signal composition as a sum of two harmonics can be written as

$$x(t) = A_1\cos\omega_1 t + A_2\cos\left(\omega_2 t + \varphi\right) \tag{15}$$

where φ is the initial phase shift angle. The corresponding first derivative of the signal is

$$\dot{x}(t) = -A_1\omega_1\sin\omega_1 t - A_2\omega_2\sin\left(\omega_2 t + \varphi\right) \tag{16}$$

Every zero-crossing of the first derivative corresponds to the existence of a local extremum of the initial function. At a certain moment t_i, as the first derivative is equal to zero a single local extremum, for example, maximum $x_{\max}(t_i)$ occurs. The closest single local minimum $x_{\min}(t_j)$ occurs at another certain moment t_j, so every closest maximum and minimum exist at different moments ($t_i \neq t_j$). However, the EMD method requires that both the top and the bottom extremum curves be constructed of synchronous moments [1] . For every top maximum we need to construct its virtual synchronous bottom pair, and correspondingly for every bottom minimum its virtual synchronous top pair. Let us analyze two closest neighbor extrema. The original EMD method as known uses interpolation with cubic spline to build the synchronous top and bottom line [1].

To simplify, for each maximum we find the closest minimum from the left and the closest minimum point from the right. Then we estimate the mean (median) value of these two neighboring minimum points before and after the maximum, thus yielding the desired virtual synchronous bottom pair of each maximum. By ending up with all virtual synchronous pairs, we obtain the bottom extremum line required by the EMD. The

proposed simplest short straight line length fitting makes it possible to analyze and understand the main properties of the EMD.

By analogy, the mean value of two neighboring maximum points before and after the minimum will produce the desired virtual synchronous top pair of the minimum. As a result, for the initial signal composition with two sets of maxima and minima we will get two corresponding synchronous top and bottom lines constructed from short straight line lengths [7] .

Next, the EMD requires computing the arithmetic mean value of the obtained top and bottom lines [1]. As shown, the extrema wander throughout the signal values, so the mean value will depend on the current position of the local extrema (Eq.(9)). Let us describe two different extreme cases: a) the highest current position of the local maximum when $t_i = 0, \varphi = 0$, and b) the lowest current position of the local maximum when $t_i = \pi, \varphi = \pi$. All other middle positions of the local maximum between these two extreme cases will exhibit only intermediate behavior.

4.5.1 The case of $t_i = 0, \varphi = 0$

The initial maximum value is $x_{\max}(0) = A_1 + A_2$ and the initial value of the first derivative is $\dot{x}(0) = 0$. The closest minimum value corresponds to the next zero value of the first derivative $\dot{x}(\Delta t) = A_1\omega_1 \sin\omega_1\Delta t + A_2\omega_2 \sin(\omega_2\Delta t) = 0$. The last nonlinear equation can be solve analytically by $\Delta t = -\omega_1^{-1}\arcsin\left[\frac{A_2\omega_2}{A_1\omega_1}\sin(\omega_2\Delta t)\right]$, if $A_1\omega_1 \geq A_2\omega_2$, or $\Delta t = -\omega_2^{-1}\arcsin\left[\frac{A_1\omega_1}{A_2\omega_2}\sin(\omega_1\Delta t)\right]$, if $A_1\omega_1 < A_2\omega_2$. The obtained solution of the time moment for the closest minimum value Δt depends only on the amplitude and frequency ratios of the harmonics A_2/A_1 and ω_2/ω_1 .

This obtained solution makes it possible to generate the closest minimum values from the left and from the right, and since the cosine is the even function ($x_{\min}(\Delta t) = x_{\min}(-\Delta t)$), the virtual synchronous minimum value will be: $x_{\min}(0) = A_1\cos\omega_1\Delta t + A_2\cos(\omega_2\Delta t)$. The arithmetic average of the initial maximum and the obtained synchronous minimum can be written in the form $\mathrm{F}_1 = 0.5\left[x_{\max}(0) + x_{\min}(0)\right] = 0.5\left[A_1 + A_2 + A_1\cos\omega_1\Delta t + A_2\cos(\omega_2\Delta t)\right]$.

The obtained solution is convenient to divide and analyze separately for two parts, with the first showing only the first harmonics modification (Figure 2, a)

$$\mathrm{F}_{1,1} = 0.5A_1\left[1 + \cos\omega_1\Delta t\right] \tag{17}$$

and the second part showing the second harmonics modification (Figure 2, b)

$$\mathrm{F}_{1,2} = 0.5A_2\left[1 + \cos(\omega_2\Delta t)\right]. \tag{18}$$

Each of these parts describes value of the arithmetic mean between the top and bottom extrema in the highest current position of the local maximum when $t_i = 0, \varphi = 0$.

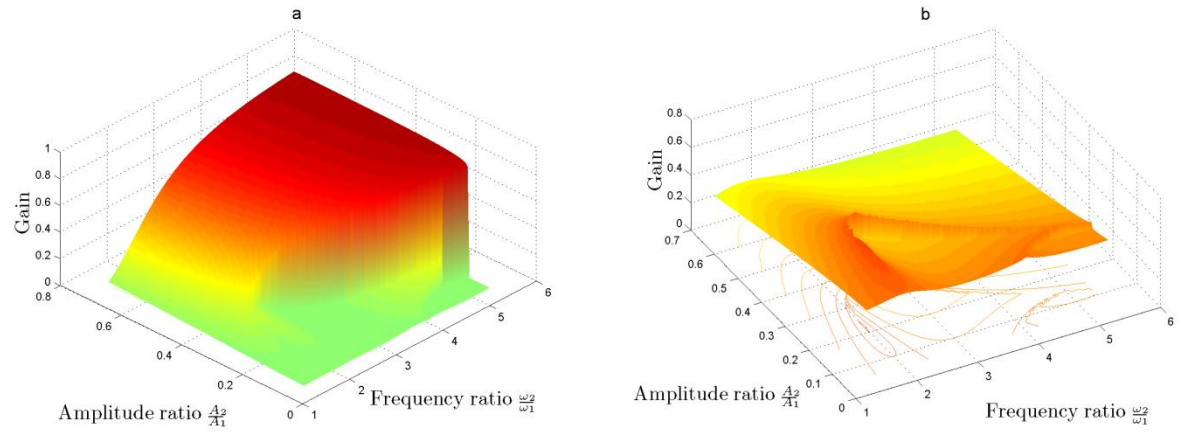

Figure 2.

Theoretical mean value between the local maxima and minima at the highest maximum position: (a) the envelope of the first harmonics, (b) the envelope of the second harmonics.

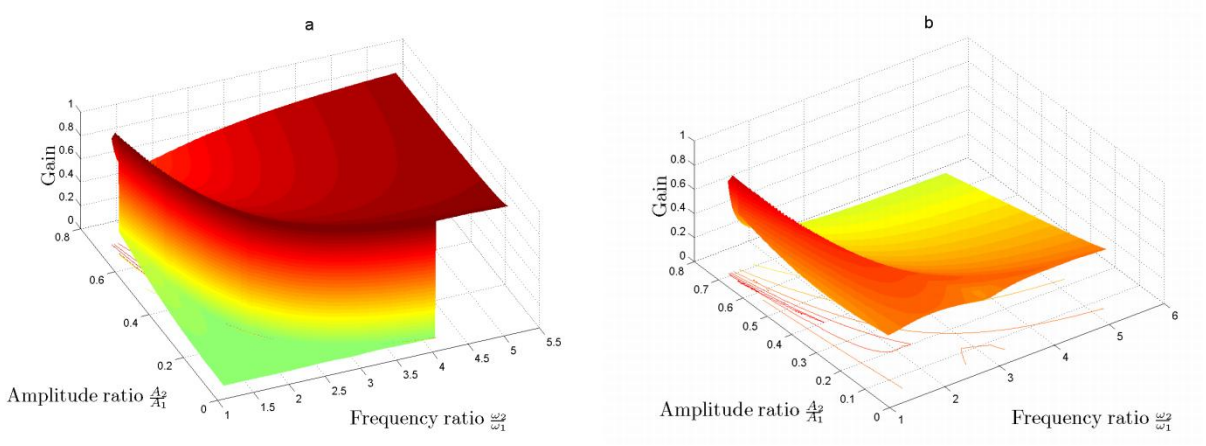

Figure 3.

Theoretical mean value between the local maxima and minima at the lowest maximum position: (a) the envelope of the first harmonics, (b) the envelope of the second harmonics.

4.5.2 The case of $t_i = \pi, \varphi = \pi$

The initial maximum value is $x_{\max}(\pi) = -A_1 + A_2$ and the initial value of the first derivative is $\dot{x}(\pi) = 0$. The closest minimum value corresponds to the next zero value of the first derivative $\dot{x}(\Delta t) = -A_1\omega_1 \sin\omega_1\Delta t + A_2\omega_2 \sin(\omega_2\Delta t) = 0$. The solution of the last nonlinear equation determines the time moment for the closest minimum value Δt, which depends only on the amplitude and the frequency ratios of the harmonics A_2/A_1 and ω_2/ω_1.

The obtained solution allow generation of the closest minimum values from the left and from the right, and since the cosine is the even function ($x_{\min}(\Delta t) = x_{\min}(-\Delta t)$), the virtual synchronous minimum value will be: $x_{\min}(\pi) = -A_1 \cos\omega_1\Delta t + A_2 \cos(\omega_2\Delta t)$. The arithmetic average of the initial maximum and the obtained

synchronous minimum takes the form

$$F_2 = 0.5\left[x_{\max}(\pi) + x_{\min}(\pi)\right] = 0.5\left[-A_1 + A_2 - A_1\cos\omega_1\Delta t + A_2\cos\left(\omega_2\Delta t\right)\right].$$

Again the obtained solution is convenient to divide and analyze separately in two parts, with the first showing only modification of the amplitude of the first harmonics (Figure 3, a)

$$F_{2,1} = \left|-0.5A_1\left(1+\cos\omega_1\Delta t\right)\right| \tag{19}$$

and the second showing only modification of the amplitude of the second harmonics (Figure 3,b)

$$F_{2,2} = 0.5A_2\left[1+\cos\left(\omega_2\Delta t\right)\right]. \tag{20}$$

Each part describes the value of the arithmetic mean between the top and bottom extrema in the lowest current position of the local maximum when $t_i = \pi, \varphi = \pi$.

4.6 EMD as a nonstationary and nonlinear filter

The final step of the EMD algorithm subtracts the obtained arithmetic mean from the initial signal [1] . Such subtraction gives rise to the "intrinsic mode function" (IMF). Therefore, the subtraction specifies a digital filtering operation where the input is the initial signal composition; the output is the IMF, and the filter characteristics (Eq.(17)-(19)) are represented in Figure 2 and Figure 3.

In order to understand how the IMF is extracted from the initial signal, let us analyze the obtained filter characteristics. This analytical three-dimensional filter is defined as a 3D function with two arguments: the relative harmonics amplitude ratio A_2/A_1 and the relative harmonics frequency ratio ω_2/ω_1. The vertical gain value of the surface presents a portion of the magnitude passing through the filter. Vertical gain values that are close to 1 indicate that the output signal passes through the filter, while those that are close to 0 indicate rejection of the output signal. The obtained analytical solutions demonstrate that the EMD is nonstationary and nonlinear at the final step of estimating the IMF. Being nonstationary means that the filter's characteristics are varied, with the extremum roaming from the top to the bottom position. The nonlinearity is embedded in the filter magnitude's dependency on the harmonics amplitude and frequency ratio.

The main common property of the obtained analytical filters for both the top and the bottom extrema positions is their high gain magnitude values for the first harmonics when $\omega_2 \to \infty$ and $A_2 \to 1$. This is a tendency of the high pass frequency to pass the unmodified first harmonics with larger frequency and amplitude ratio.

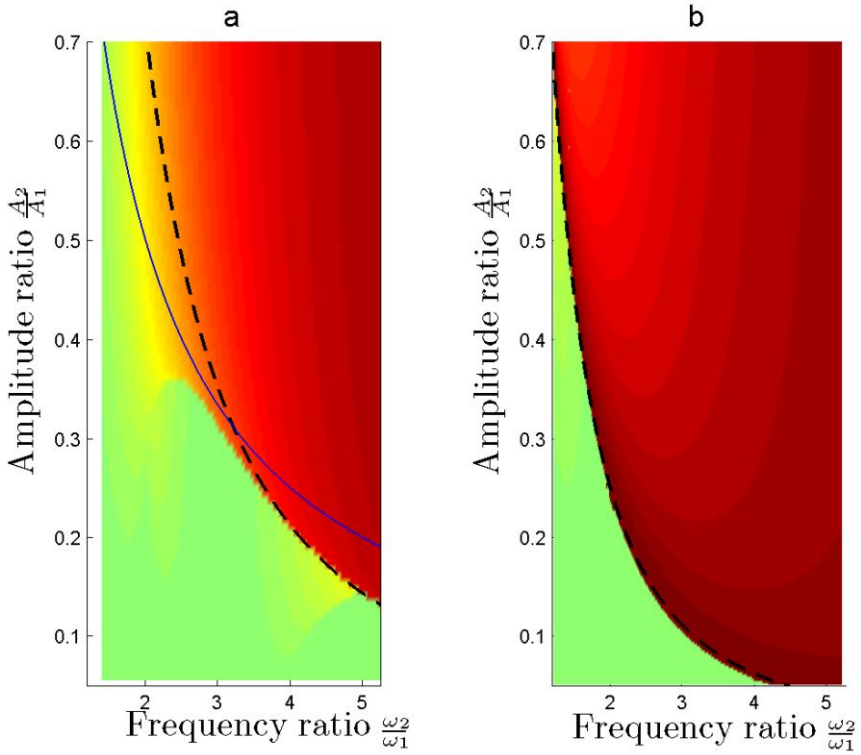

Figure 4.

Theoretical boundary of the first harmonics filtering: (a) the highest maximum position, the approximation: $A_2/A_1 \le 2.4(\omega_2/\omega_1)^{-1.75}$ (--), the approximation $A_2/A_1 = \omega_1/\omega_2$ (–); (b) the lowest maximum position, the approximation $A_2/A_1 \le (\omega_2/\omega_1)^{-2}$ (--).

Another common property of the analytical filters is their high pass magnitude values for the first harmonics and small (practically rejection) magnitude values for the second harmonics. As a result, the first harmonics with low frequency $A_1 \cos \omega_1 t$ can fully pass through the filter, while the second harmonics with high frequency will be stopped. Then, after being subtracted from the initial composition, the low frequency harmonics will disappeared, and the final IMF will consist only of the second high frequency harmonics $A_2 \cos \omega_2 t$. One more important common property of the analytical filters is the existence of a separation boundary surface $\mathrm{B}(A_2/A_1,\ \omega_2/\omega_1)$ dividing the space of parameters for two ranges and thus allowing the low frequency harmonics through or not (Figure 4.).

4.7 Frequency resolution of the EMD

The revealed separation boundaries are related directly to the frequency resolution characteristics of the EMD. For more precise analysis, let us plot the 2D projection of the same nonlinear filters with the axes A_2/A_1 and ω_2/ω_1 as pseudocolor graphs (Figure 4). Changes in such 2D graph intensities are usually defined much more clearly. Depending on the amplitude and frequency ratio, the limiting boundary determines the region to the right where the EMD is able to separate harmonics and the region to the left where the EMD cannot separate two tones.

The top extrema position filter (Figure 4, a) has a soft slope and a cutoff boundary that runs in the direction of higher frequency from the right side. A good power approximation of the boundary (Figure 4, a, dashed line) shows that the filter blocks the first harmonics when $\left(A_2/A_1\right)_{\text{boundary(top)}} \le 2.4\left(\omega_2/\omega_1\right)^{-1.75}$, and it passes the first harmonics without modification for higher relations $A_2/A_1 > \left(A_2/A_1\right)_{\text{boundary(top)}}$. In the same figure (Figure 4, a, thick line) we plotted the curve corresponding to the inversely related amplitude and frequency ratio $A_2/A_1 = \omega_1/\omega_2$ for the case of negative instantaneous frequency (Eq.(12)). This inversely related amplitude and frequency ratio is rather close to the theoretical boundary curve.

The bottom extrema position filter (Figure 4, b) has a hard slope. Its cutoff boundary is shifted to the left, and the filter blocks the first harmonics exactly when $\left(A_2/A_1\right)_{\text{boundary(bottom)}} \le \left(\omega_2/\omega_1\right)^{-2}$. The harmonics ratios located from the left to the boundary of the hard bottom filter does not allow the EMD to extract the harmonics at all despite any large number of sifting iterations See an attempt of the EMD decomposition of very frequency close harmonics in Figure 5. It is evident that the harmonics ratios located from the right to the boundary of the soft top filter let the EMD extract the harmonics completely and at once during the first iteration. An example of the complete decomposition of the distant harmonics is presented in Figure 6.

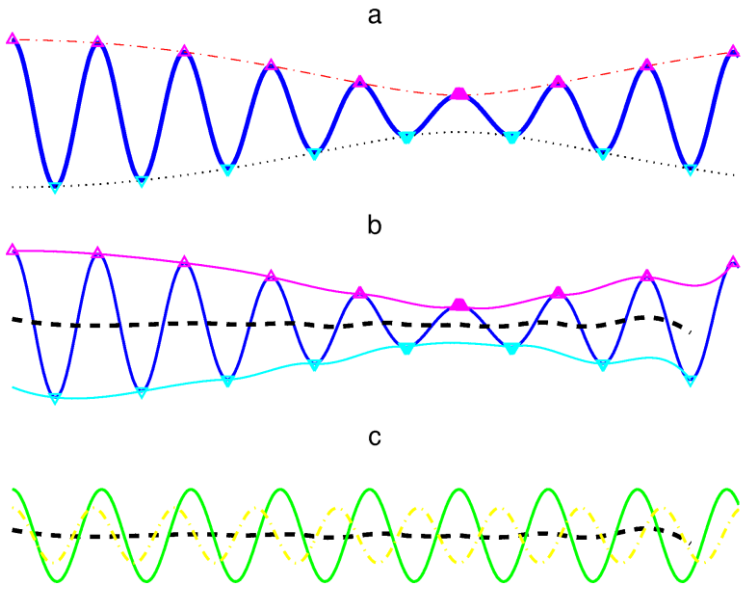

Figure 5.

Composition of two very close harmonics ($A_1 = 1$, $\omega_1 = 1$, $A_2 = 0.6$, $\omega_2 = 1.1$): (a) the initial signal (▬), the upper (-·-) and the lower (···) envelope, the top (Δ) and the bottom maxima points (▽); (b) the initial signal (▬), the top (–) and the bottom (–) maxima curves, the mean value between the top and the bottom maxima curves (--); (c) the first harmonics (–), the second harmonics (-·-), the mean value between the top and the bottom maxima curves (--).

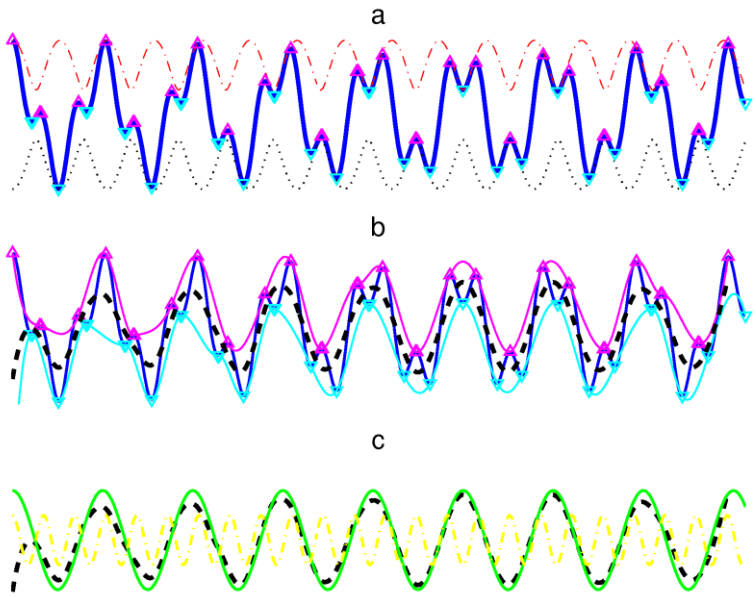

Figure 6.

Complete decomposition of two distant harmonics ($A_1 = 1, \omega_1 = 1, A_2 = 0.4, \omega_2 = 2.9$): (a) the initial signal (▬), the upper (-·-) and the lower (···) envelope, the top (Δ) and the bottom maxima points (▽); (b) the initial signal (▬), the top (—) and the bottom (—) maxima curves, the mean value between the top and the bottom maxima curves (**--**); (c) the first harmonics (—), the second harmonics (-·-), the mean value between the top and the bottom maxima curves (**--**).

In the case when the frequency of the harmonics intervenes between these boundaries (Figure 4), the first harmonics at every sifting iteration will pass the filter partially, with the attenuation coefficient depending on the filter gain (see Figure 7.). To approach the full value of the envelope, the first harmonics should be passed through the filter several times. In other words, harmonics whose frequency ratio is located between these theoretical boundaries might be separated during several sifting iterations.

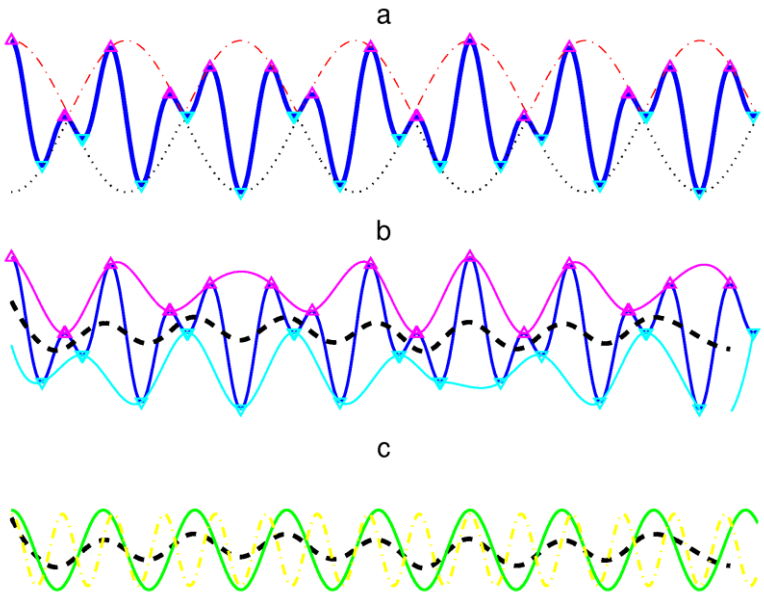

Figure 7.

Partial decomposition of two close harmonics ($A_1 = 1, \omega_1 = 1, A_2 = 0.9, \omega_2 = 1.8$): (a) the initial signal (▬), the upper (-·-) and the lower (···) envelope, the top (Δ) and the bottom maxima points (▽); (b) the initial signal (▬), the top (—) and the bottom (—) maxima curves, the mean value between the top and the bottom maxima curves (**--**); (c) the first harmonics (—), the second harmonics (-·-), the mean value between the top and the bottom maxima curves (**--**).

Logically the presented analytical nonstationary and nonlinear filters describe the EMD separation capacity only in the extreme positions. In reality, the current positions of the extrema change constantly and the filters are continuously being transformed from one to another, producing a kind of mid-position filtering. The spline fitting used in the real EMD program also can have some impact on the filter's slope, but both obtained extreme analytical boundaries will remain unchanged.

Thus, the more frequencies are spaced apart, the smaller amplitude ratio of two harmonics is suitable for EMD separation. For example, a second harmonics with a tripled or lower frequency $\omega_2 \le 3\omega_1$ and a small amplitude less than $A_2 < 0.3A_1$ can be extracted with some iterations. Nevertheless, a smaller amplitude that is less than $A_2 \le 0.11A_1$ absolutely cannot be separated by the EMD. For example, if frequencies lie within an octave of each other $f_2 \le 2f_1$ and their amplitudes differ less than $A_2 \le 0.25A_1$, the EMD method practically is unable to separate such two components. This means that the EMD does not perform well for smaller amplitudes of the second harmonics and cannot distinguish frequencies that are close to each other. For these close frequencies is more suitable the Hilbert Vibration Decomposing (HVD) method which has much better frequency resolution for adaptive decomposition of nonstationary and AM modulated signals [6-7].

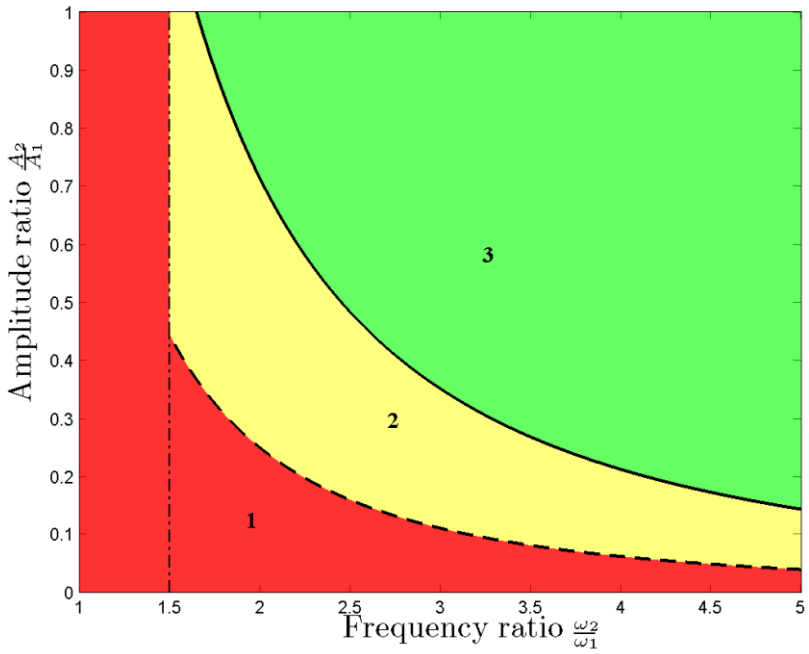

Figure 8

Summary of the EMD ranges for harmonics separation: (1) impossible decomposition for very close frequency harmonics and small amplitude ratio; (2) decomposition requires several sifting iterations for close frequency harmonics; (3) single iteration separation for distant frequency harmonics and large amplitude ratio.

The logical result of the provided theoretical analysis is that the frequency and amplitude ratios of the harmonics can be separated into three different groups (Figure 8): (1) harmonics with very close frequencies and small amplitude $A_2/A_1 \le (\omega_1/\omega_2)^2$ unsuitable for EMD decomposition, (2) close frequency harmonics

$(\omega_1/\omega_2)^2 \le A_2/A_1 < 2.4(\omega_1/\omega_2)^{1.75}$ requiring several sifting iterations, and (3) distant frequencies and large amplitude harmonics $A_2/A_1 \le 2.4(\omega_1/\omega_2)^{1.75}$ that are well separated for a single iteration.

4.8 The EMD frequency limit undependable of the amplitude ratio

According to Eq.(14), the continuous function of the cosine projection $x_{extr}(t)$ oscillates with the frequency $\omega_2 - \omega_1$ posed by the envelope oscillation. This function is sampled with a sampling frequency equal to the instantaneous frequency $\omega(t)$ to form the extrema points. If the frequency of the oscillation is larger than the Nyquist frequency $\omega_2 - \omega_1 > 0.5\omega(t)$, the sampled extrema curves undergo nonlinear filtering, as described in the previous sections. If the frequency of the oscillation is less than the Nyquist frequency

$$\omega_2 - \omega_1 \le 0.5\omega(t), \tag{21}$$

no aliasing or folding occurs and the resultant sampled extrema curve will oscillate with the same initial envelope frequency $\omega_2 - \omega_1$. Such retention of the frequency means that the top extrema will repeat the upper envelope and correspondingly the bottom extrema curve will repeat the lower envelope. As a result, an averaging of theses extrema curves always will produce zero and the EMD will not be able to decompose harmonics.

The simple formula in Eq.(21) provides a strong limit of operationability for the EMD method undependable of the amplitude harmonics ratio. As shown in Eq.(13), the average value of the instantaneous frequency of the composition is equal to the frequency of the largest harmonics. In our notations the average frequency of two harmonics always is $\overline{\omega}(t) = \omega_1$. Substituting this value in Eq. (21) yields $\omega_2 - \omega_1 \le 0.5\omega_1$ or $\omega_2 \le 0.5\omega_1 + \omega_1$ and finally

$$\omega_2 \le \frac{3}{2}\omega_1 \tag{22}$$

Eq. (22) yields the smallest value of second harmonics frequency that the EMD is able to distinguish. If the value of ω_2 is any lower, the EMD is unable to distinguish the components. The obtained smallest value of $\omega_2/\omega_1 = 1.5$, shown in Figure 8, is an absolute strong limit that does not depend on the amplitude relations between harmonics. This theoretical limit value completely coincides with the experimental critical frequency ratio $\omega_1/\omega_2 \approx 0.67$ found in [5] . Above this value, it is impossible to separate the two components no matter what the amplitude ratio. This is the case when the local extrema do not differ from the corresponding envelope curves.

Conclusion

Using the first derivative of the signal in the signal analytic form, we devised an expression for the local extremum points, including their vertical locations and distribution in time. As shown above, the obtained vertical position of the local extrema can deviate from the envelope, thus explaining, for instance, why and when the maximum points can become negative.

The revealed extrema deviation from the envelope forms the basis for a theoretical explanation of the EMD sifting procedure. It was shown that the vertical distance between the envelope and the local extrema depends on the relation between the first derivative of the envelope from one side and the multiplication of the envelope by the instantaneous frequency from the other.

To build the simplest synchronous extrema, we suggested connecting the opposite closest neighboring left and right extrema, thus yielding a theoretical median function between the top and bottom extrema of two harmonics.

This theoretical median function represents a kind of signal nonlinear filter whose input is an initial two-tone composition and whose output can be the harmonics with the lowest frequency. Depending on the harmonics amplitude and frequency ratios, the filter passes through some portion of the magnitude of the lowest frequency. The filter is nonstationary because its characteristics vary, while the extrema roam from the highest to the lowest position. At these extreme positions, the filter characteristics differ: at the highest position, the filter has a soft slope, but at the lowest position it has a hard slope.

The obtained boundaries between the filter pass and the stop characteristics determine the theoretical frequency resolution of the EMD. When the amplitude of the smaller harmonics is less than the boundary of the hard slope, the EMD does not separate the harmonics. When the amplitude of the smaller harmonics is larger than the boundary of the soft slope, the EMD separates the harmonics according to its first single sifting iteration. Middle amplitudes between the boundaries require several iterations, depending on the filter attenuation pass characteristics. In such a manner the initial composition after extraction of the median function will contain a high frequency harmonics such as the intrinsic mode function. This explains how the EMD used sifting to decompose the first high frequency components.

For two-tone models, the critical frequency limit of distinguishing the closest harmonics was found theoretically. The harmonics with a frequency below the critical frequency cannot be extracted by the EMD from the decomposition, no matter how large its amplitude.

Like any other signal processing procedure, the EMD operates with an input signal only. The EMD decomposes the signal exclusively by means of its inherent geometric transformation function. For a composition of harmonics it extracts at first the highest frequency of the composition. Like any other signal analysis instrument, it merely reflects and represents real physical and natural processes and phenomena. It

is not feasible to try to understand or explain the EMD tool through nonlinear structural dynamics or through any other physics-based foundation.

References

[1] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, et al. The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis. *Proc. R. Soc. London*, Ser. A (1998) 454, pp. 903–995.

[2] Z. Wu and N. E. Huang. A study of the characteristics of white noise using the empirical mode decomposition method, *Proc. R. Soc. Lond. A*, 2004, Vol. 460, pp. 1597–1611.

[3] Wu, Z., and N. E Huang (2008), Ensemble Empirical Mode Decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis. Vol.1*, *No.1*. 1-41.

[4] S. Kizhner, K. Blank, T. Flatley, N. E. Huang, D. Petrick, Ph. Hestnes. On Certain Theoretical Developments Underlying the Hilbert-Huang Transform. *Aerospace Conference*, 2006 IEEE, March 2006, 14 pp.

[5] G. Rilling and P. Flandrin. One or Two Frequencies? The Empirical Mode Decomposition Answers. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 2008, V. 56, Issue: 1, pp. 85-95.

[6] M. Feldman M. Time-Varying Vibration Decomposition and Analysis Based on the Hilbert Transform. Journal *of Sound and Vibration*. 2006, Vol 295/3-5 pp. 518-530.

[7] M. Feldman. Theoretical analysis and comparison of the Hilbert transform decomposition methods, *Mechanical Systems and Signal Processing*, 2008, Volume 22, Issue 3, pp. 509-519.

[8] M. Feldman. Analytical Basics of the EMD: Two Harmonics Decomposition. *Mechanical Systems and Signal Processing*, 2009, Volume , Issue , pp. - ,