

On the Hilbert-Huang Transform Theoretical Developments

Semion Kizhner, Karin Blank, Thomas Flatley, Norden E. Huang, David Patrick and Phyllis Hestnes

*National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt Road, Greenbelt MD, 20771
301-286-1294
Semion.Kizhner-1@nasa.gov*

Abstract

One of the main heritage tools used in scientific and engineering data spectrum analysis is the Fourier Integral Transform and its high performance digital equivalent – the Fast Fourier Transform (FFT). Both carry strong a-priori assumptions about the source data, such as linearity, of being stationary, and of satisfying the Dirichlet conditions. A recent development at the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC), known as the Hilbert-Huang Transform (HHT), proposes a novel approach to the solution for the nonlinear class of spectrum analysis problems. Using a-posteriori data processing based on the Empirical Mode Decomposition (EMD) sifting process (algorithm), followed by the normalized Hilbert Transform of the decomposition data, the HHT allows spectrum analysis of nonlinear and nonstationary data. The EMD sifting process results in a non-constrained decomposition of a source real value data vector into a finite set of Intrinsic Mode Functions (IMF). These functions form a near orthogonal adaptive basis, a basis that is derived from the data. The IMFs can be further analyzed for spectrum interpretation by the classical Hilbert Transform. A new engineering spectrum analysis tool using HHT has been developed at NASA GSFC, the HHT Data Processing System (HHT-DPS). As the HHT-DPS has been successfully used and commercialized, new applications post additional questions about the theoretical basis behind the HHT and EMD algorithms. Why is the fastest changing component of a composite signal being sifted out first in the EMD sifting process? Why does the EMD sifting process seemingly converge and why does it converge rapidly? Does an IMF have a distinctive structure? Why are the IMFs near orthogonal? We address these questions and develop the initial theoretical background for the HHT. This will contribute to the developments of new HHT processing options, such as real-time and 2-D processing using Field Programmable Array (FPGA) computational resources, enhanced HHT synthesis, and broaden the scope of HHT applications for signal processing.

1. Introduction and Research Methodology

1.1 Introduction

One of the main heritage tools used in scientific and engineering data spectrum analysis is the Fast Fourier Transform that carries strong a-priori assumptions about the source data, such as linearity and of being stationary. The data must originate from a periodic waveform and also satisfy the Dirichlet conditions of having a finite number of discontinuities and extremas, and be integrable in any sequence of time intervals with length of the period T [8]. Heritage spectrum analysis methods use a fixed basis in their transforms while EMD derives its basis adaptively from the data itself.

The EMD sifting process [1]-[3] is a novel algorithm for digital signal processing of non-linear and nonstationary data. Given an arbitrary input vector of rational numbers, the EMD algorithm invariably sifts out IMF components of different time scales with the fastest varying component being sifted out first. This is accomplished by a process known as “sifting” which is repeatedly applied to the signal until it converges on criteria that defines an IMF. In addition, the EMD process itself converges and produces a finite number of IMFs. All this has been observed in all tests conducted so far, in other words, empirically.

However, it is not that obvious why the EMD algorithm behaves this way and this paper addresses these questions by examining a few analogies and intuitive examples of signals containing artificially created fast and slow varying components. We also consider a general case for the EMD sifting process and establish the theoretical foundations of the EMD algorithm’s sifting sequence of scales and theoretical convergence. We accomplish this by providing a hypothesis for why the fastest scale is sifted out first and reporting two hypotheses about the sifting process’s symmetric pair invariance and resulting IMF structure, and its theoretical rapid convergence in diminishing amplitude regions. We then propose a few new applications based on this research, as well as define the areas of future research work on the HHT.

2. EMD Algorithm Overview and Problem Statement

The EMD algorithm’s empirical behavior is determined by its built-in definitions and criterias as well as by the user’s supplied run configuration vector Ψ as described in more details in Section 2.1

$$\Psi = \Psi(\Delta t, m, k, p, \dots) \quad (1)$$

Ψ is composed of empirical parameters supplied by a user when running HHT-DPS. Among these parameters are Δt , the sampling time interval, m is the maximum number of allowable IMFs to strive for, k is the maximum allowable number of EMD sifts for one IMF, and p , which allows user to select the “beyond-the-envelope-end-points” prediction algorithm option. The following formulation of the Empirical Mode Decomposition algorithm is describing the EMD implementation in the HHT-DPS Release 1.4.

2.1 HHT-DPS EMD Algorithm

The EMD algorithm is based on a few distinct procedural steps (a) - (g) as follows:

EMD Algorithm Entry Point

User specifies EMD Configuration Vector $\Psi = \Psi(\Delta t, m, k, p, \dots)$ and input signal $s(t)$, where s is the energy values and t is the equally spaced time values t_i for $i \in [1, \dots, N]$. In the first iteration of this algorithm, the “data parent” is the original signal.

EMD Sifting Process Iterative Loop Entry Point

(a) The first step of the algorithm is to search the parent dataset and find the local maximum and minimum values, also known as extrema. These points are divided into two sets, p^{\max} and p^{\min} , containing the respective local maximum and minimum extrema values.

(b) The two subsets for both extrema point types p^{\max} and p^{\min} are splined using *piecewise cubic splines*, comprising the upper and lower boundary of the extrema subsets' envelope boundaries $u(t)$ and $d(t)$. *The use of the cubic spline provides a relatively slow changing background median $M(t)$ cubic spline* (See (d) below) against which local fast variations of the signal become prominent and which, in turn, forms the basis for the most variant signal being sifted out first by the EMD sifting process. But first the extrema sets are extended a bit beyond $[t_1, t_N]$ to always maintain data over the original argument interval, as described next in (c).

(c) The sets $u(t)$, $d(t)$ are extended a bit beyond the data original argument interval $[t_1, t_N]$ with a few predicted maxima and minima extrema points on both ends of the original argument interval $[t_1, t_N]$. There are many prediction algorithms, meaning there is no accepted one. The extended extrema sets are then splined beyond original argument interval $[t_1, t_N]$ using predicted extrema points and then re-sampled on $[t_1, t_N]$. This original argument interval is always maintained throughout the HHT-DPS.

(d) As stated above in (c), splines $u(t)$, $d(t)$ are then re-sampled for sampling time arguments t_i , where $i \in [1, \dots, N]$, and the discrete median vector $M(t_i)$ is computed as

$$M(t_i) = (u(t_i) + d(t_i)) / 2 \quad (2)$$

This definition of median is essential, because it assures the sifting process convergence. When the median computation is iterated k -times in the sifting process, the divisor becomes 2^k , which rapidly becomes a large number. The convergence theoretics are based on this number and are described in more details in Sections 3.

(e) The difference $r(t_i) = p(t_i) - M(t_i)$ is formed and the *sifting residue* $r(t_i)$ is checked against the IMF criteria, as follows in (f).

(f) The IMF Criteria

The difference between the number $E(r(t_i))$ of extrema points in $r(t_i)$, and the number $Z(r(t_i))$ of its zero-crossings or the change in two adjacent extrema points' magnitude sign, may not vary by more than 1. Additionally, each IMF must have at least 3 extrema points

$$(E(r(t_i)) > 3) \ \&\& \ (|E(r(t_i)) - Z(r(t_i))| \leq 1), \quad i \in [1, \dots, N] \quad (3)$$

An $r(t_i)$, which satisfies the IMF Criteria, is called an IMF. It is stored during the EMD algorithm verification of the IMF Criteria and it is then subtracted from the current signal residue $R(t)$ to obtain the EMD next sifting parent signal $p(t)$. This signal residue becomes the new input parent to the EMD sifting process iteration step. *Note, that at no time does a median $M(t)$ become an input to the EMD sifting process.*

The set of IMFs, which is derived from the data, comprises the signal $s(t)$ near-orthogonal adaptive basis and is used for the following signal time-spectrum analysis.

If the sifting residue $r(t_i)$ satisfies the IMF Criteria the IMF is retained and control is passed to the Entry Point (a).

(g) Process Completion Criteria

When $r(t_i)$ is checked above in (e) against the IMF Criteria and Exit Criteria

$$((E(r(t_i)) \leq 3) \text{ OR } (\#IMFs = m))$$

The EMD sifting process completes with the last signal residue $R(t_m)$ becoming the process residue from which an IMF couldn't be made.

The EMD process completes and exits.

If $r(t_i)$ does not satisfy the Completion Criteria, the control is passed to the **Entry Point, a).**

Naturally, because of the EMD algorithm construct, the sum of all IMF_k and the last signal residue $R(t_m)$ synthesize the original input signal $s(t)$

$$s(t) = \sum IMF_j + R_m, \text{ where } 1 \leq j \leq m-1 \quad (4)$$

Additional details concerning the EMD can be found in references [4]-[5].

2.2 Problem Statement

With the EMD algorithm described above, the research problem is to understand why it works this way and try to develop the theoretical fundamentals of this algorithm.

3. Hypothesis 1 and Theory of the EMD Sifting Process Sequence of Scales

The fastest changing component of a composite signal is invariably *being sifted out first* in the Empirical Mode Decomposition algorithm and the first hypothesis in the development of the EMD algorithm theory is related to the question: "why?" This empirical fact is significant in the respect that it provides the first insight into how the EMD algorithm works. Also the EMD algorithm's computational performance depends on the number of extrema in the input data set and taking out first all the extrema for the fastest varying component greatly contributes to the EMD algorithm's next IMF sifting steps computational performance.

Hypothesis 1: *Assuming theoretical convergence of the EMD sifting process, the fastest scale is being sifted out first, because the composite signal $s(t)$ extremas' envelope median is approximating the slower variance signal in presence of a fast varying component.*

It is implied in this Section that the EMD sifting process converges theoretically in all cases. The EMD sifting process rapid theoretical convergence hypotheses are formulated in this paper and will be reported in future papers.

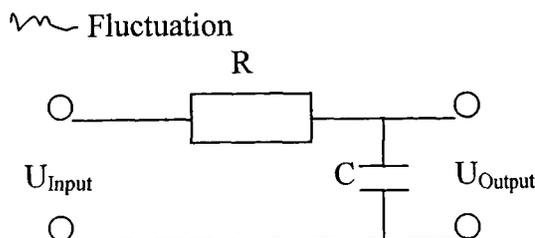
In order to prove this hypothesis we are first considering *two analogies to the EMD sifting process* - one from optical physics and the other from electrical and electronics engineering disciplines. These analogies provide initial insight into why the signal fastest changing component is being sifted out first by the EMD sifting process. We then consider *three examples* of the EMD sifting for a few artificially created signals comprising of fast and slow varying components. These analogies and intuitive examples allow us to gain an initial insight into the mechanism of why the fastest scale gets sifted out first. We then examine the EMD sifting of an arbitrary signal and prove the hypothesis for a general case.

3.1 Analogy 1, Light Spectrum

The first analogy has to do with light. Light spectrum analysis, which has played such an important part in astronomical work, is essentially a method of ascertaining the nature of a remote celestial body by *a process of sifting or analyzing into different components the light received from them* [9]. It was first clearly established by Newton that ordinary *white light is comprised of waves of different frequencies*. The prism sifts out different colors. This known empirical phenomenon has a well-established *theoretical basis* in that the prism bends light of different wavelength by a different degree ρ , *resulting in a distinctive order of output light spectrum (colors), with higher frequency components appearing "first" in the prism spectrum. It results in a definitive sequence of colors at the output of a prism.*

3.2 Analogy 2, RC-Chain Filter

The second analogy has to do with and electrical RC-chain (circuitry). This circuitry comprises a resistance (R) and capacitor (C), with the RC-chain characteristic constant $\tau = RC$ as depicted in Picture 1. This circuitry is often used to *smooth (filter out) fast variable fluctuations of a constant input voltage U_{Input}* . The functioning of the RC-chain as a filter depends on its user selected hardware parameter, the characteristic constant τ .



Picture 1. RC-chain

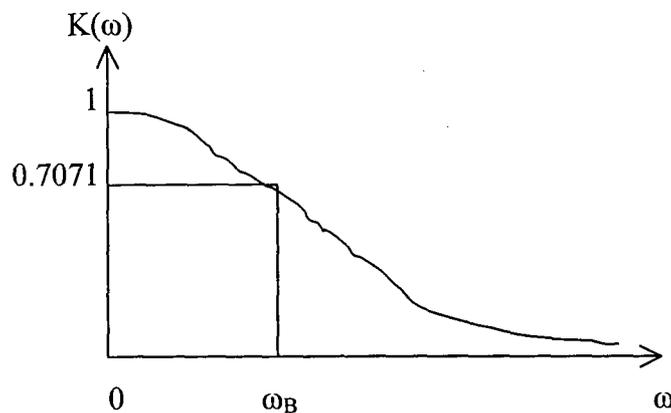
The RC-chain's transfer function $K(\omega)$ [7] can be described as the ratio of the output voltage to input voltage, with both voltages considered as a function of frequency ω :

$$K(\omega) = U(\omega)_{\text{Output}} / U_{\text{Input}}(\omega) = 1 / \sqrt{1 + \omega^2 C^2 R^2}$$

or

$$K(\omega) = 1 / \sqrt{1 + \omega^2 \tau^2} \quad (8)$$

For a near constant input voltage ($\omega=0$, absence of fluctuations) the capacitor impedance $X_C = 1/\omega C = 1/0 \cdot C = \infty$ and $K(0) = 1$. As $\omega \rightarrow \infty$ increases, the capacitor impedance is decreasing, $X_C \rightarrow 0$, and the amplitude of the output voltage variable component factor $K(\omega)$ is approaching 0, as shown on the graph in Picture 2.



Picture 2. RC-chain Transfer Function

If we arbitrarily select the output voltage amplitude as a fraction of the input voltage, for example, as $K(\omega) = 0.707 = 1/\sqrt{2}$ (this threshold value of $1/\sqrt{2}$ is selected to get the following convenient computations and is actually the rmf power of a sinusoid voltage function), then the corresponding frequency band ω_B or filter upper boundary can be evaluated from the transfer function as follows:

$$1 / \sqrt{2} = 1 / \sqrt{1 + \omega_B^2 \tau^2}$$

or

$$(1 + \omega_B^2 \tau^2) = 2 \rightarrow \omega_B^2 \tau^2 = 1 \rightarrow \omega_B \tau = 1$$

or

$$\omega_B = 1 / \tau$$

Frequencies higher than ω_B are filtered out (sifted out) by this RC-chain filter and the stability of the output is increasing with the increase in the RC-chain's constant τ .

3.3 The Analogies Applicability

The EMD sifting process is computationally analogous to the "hardware" phenomenon of a prism sifting out white light into a distinctive sequence of components of different frequencies, from highest frequency in violet (790 Tera Hertz - Thz) to blue, cyan, green, yellow, orange and to lower frequency in red (480 Thz). If the bottom face of the prism is horizontal, the light beam entering from the bottom of the prism is not deviated. When

the light leaves the prism at the inclined face and enters the air, it is refracted, and the beam is deflected to the right; the deflection is larger at shorter, bluer, wavelengths.

The EMD sifting process is also analogous to the RC-chain that is sifting out the input voltage fast varying fluctuations of frequencies exceeding the less varying or a constant input voltage, as determined by the RC-chain characteristic constant τ .

Analogous to these two processes, with their characteristic parameters ρ and τ , the EMD sifting process is defined both by its *implementation definitions and criteria in the HHT-DPS*, as well as by its user supplied *empirical run-time configuration vector Ψ* .

The EMD sifts out the input signal's components in a definitive sequence of scales, analogous to that of a prism sifting out the light waveform into a sequence from shortest to longest wavelength, and performs it in a way similar to that of an RC-chain filtering out the frequency band around the slow signal component or the signal *intrinsic median level* (s), as further elaborated below.

3.4 The EMD Sifting Process, Analogies and First Intuitive Insights

Intuitively, the EMD sifting process can be initially *rationalized* in the following way. Visualize a composite signal comprising two components of which one is a highly variable component and the other is a relatively slow varying trend-like component. It is obvious that the fast varying component will be observable as a curve following in the shape of the slow varying component in the composite signal, similar to the multitudes of small inlets and bay fractals in the broad outline of an ocean shore, when observed from afar at sea, or from Space. In other words, the slow varying component can be interpreted as an approximate median of the composite signal's envelope built on fast varying component extrema points *with maybe a few standout extrema points (accidental large amplitude spikes) belonging to the slow varying component*. The signal extremas envelope's upper and lower boundaries are defined by the EMD HHT-DPS implementation as two piecewise cubic splines, connecting the signal's adjacent maxima or minima points correspondingly. The piecewise curves are also *smoothly* connected, meaning their slopes must be contiguous and be of the same value at juncture points. The EMD sifting process computes the composite signal envelope upper and lower boundaries' *median as the algebraic sum of the envelope upper and lower curves' data at sampling time points divided by 2*, and subtracts the resulting median from the parent signal to arrive at the next "sifted out" residue component candidate, as defined above in the EMD sifting algorithm overview (Section 2.1 equation (2)). Intuitively, the EMD essentially subtracts the slow varying component from the composite signal, yielding the signal's fastest varying component. This sifting residue is then checked against the IMF criteria and the process is repeated when necessary until the fastest varying IMF is found or *made* from the input composite signal.

Before proving the proposed Hypothesis 1 and elaborating on the theoretical details of the EMD sifting process, let us first consider *three intuitive examples*, using a few artificially constructed composite signals for which the workings and the results of the EMD sifting process are known in advance due to the way in which the input signals were constructed.

3.5 Three Intuitive Examples

The purpose of the following three examples is to demonstrate how the EMD sifting process works for data when the results of the sifting process *are known intuitively and a priori*. We know these results beforehand, because of the artificial way the input data in these examples was constructed. We know that the envelope of a fast varying sinusoid signal is obviously composed of two straight lines and its median is a straight line too. If we include in the signal a slower varying signal as a straight line the median intuitively will be approximated by this straight line. Considering then the low varying component a straight line (which approximates the envelope median), makes the composite signal sifting process' results intuitively clear, because the subtraction of the straight line median results in the fast varying sinusoid component. Following this line we constructed three composite signals with the purpose of observing how they are processed by the EMD.

3.5.1 Example 1

Consider a composite signal comprised of a known *constant signal* $s1=0.5$ (non-variant signal presented geometrically by a horizontal straight line) and a known relatively *fast* (as compared to $s1$) *varying signal* $s2$ that is a 5 Hz sinusoid with amplitude in the range $[-1.0, 1.0]$

$$s(t) = s1 + s2 = 0.5 + 1.0*\cos(2*\pi*f*t)$$

The EMD of this signal results in sifting out first the 2 Hz sinusoid followed by the straight line $s1=0.5$ residue.

3.5.2 Example 2

We consider now a composite signal comprised of a known and relatively slow varying signal $s1(t) = k*t$ representing an inclined straight line with slope k and a second known signal that is a 5 Hz sinusoid with amplitude 1.0, identical to $s2(t)$ in Example 1:

$$s(t) = 1*t + 1.0*\cos(2*\pi*5*t)$$

The EMD of this signal results in sifting out first the 5 Hz sinusoid followed by the line $s1=t$ residue.

3.5.3 Example 3

Finally, we consider the third intuitive example, a composite signal s comprising four components $s1, s2, s3$ and bias $b1$. Namely these are a 1 Hz, 2 Hz and 50 Hz sinusoids (cos function) and a constant bias with amplitude 1.0:

$$\begin{aligned} s1 &= b1 + 1.0*\cos(2*\pi*1*t) \\ s2 &= 2.0*\cos(2*\pi*2*t) \\ s3 &= 2.0*\cos(2*\pi*50*t) \\ s &= s1 + s2 + s3 \end{aligned}$$

The EMD of this signal results in sifting out first the 50 Hz sinusoid followed by the 2 Hz sinusoid and residue equal to component $s1$.

3.6 EMD Sifting Process Output Sequence of Scales for an Arbitrary Signal

In a general case, we now consider *an arbitrary composite signal* $s(t)$. We will prove the Hypothesis 1 for an arbitrary signal by demonstrating the reason why the fastest varying component in an arbitrary signal $s(t)$ is being sifted out first by the EMD sifting process. We are assuming the sifting process' theoretical convergence. The theoretical convergence of the EMD sifting process is proved in the following Section 4.

The only not-restrictive convention we make here is that we view an arbitrary signal being comprised of two components – a relatively fast or high variance component $s_h(t)$ and a slower or low variance component $s_l(t)$, similar to signals in the analogies and intuitive examples considered above in Sections 2.1-2.5

$$s(t) = s_h(t) + s_l(t) \quad (9)$$

This convention about the signal $s(t)$ structure, again, is based on the two analogies we had described above, and *is not-restrictive* in any sense. There is no mystery of what $s_h(t)$ may be. For example, *the locations of the initial set of extremas of $s(t)$ coincide, in general, with the location of the $s_h(t)$ extremas, except an occasional spike rooted in the remaining components. The set of extremas of the parent signal characterizes the fastest scale at the time of the sifting process.*

3.6.1.2 Linear Approximation of a Slow Varying Component

In order to explain why the fastest scale is being sifted our first priority is to consider the approximation *of the slow varying signal component $s_l(t)$ on a small interval $t \in [t_1, t_1 + \delta t]$ by a straight line passing through the interval two end points (linear approximation is here actually the definition of a slow varying component)*

$$\{ (t_1, s_l(t_1)), (t_1 + \delta t, s_l(t_1 + \delta t)) \}$$

or

$$(s_l(t) - s_l(t_1)) / (t - t_1) = (s_l(t_1 + \delta t) - s_l(t_1)) / \delta t = k$$

or

$$s_l(t) \approx (s_l(t_1 + \delta t) - s_l(t_1)) / \delta t = s_l(t_1) + k^*(t - t_1)$$

or

$$s_l(t) \approx s_l(t_1) + k^*(t - t_1)$$

Slope k is negligibly small because of the selection of $s_l(t)$ as the slow varying component and thus the straight line is being almost parallel to the Ot axis. A fast varying oscillation component combined with a line segment must satisfy above equality (9).

This leads to a median point $M(t)$ for any $t \in [t_1, t_1 + \delta t]$ being computed as

$$M(t) = ((s_l(t_1) + k^*(t - t_1)) + (\max)s_h(t)) + ((s_l(t_1) + k^*(t - t_1)) + (\min)s_h(t))) / 2 =$$

$$s_l(t_1) + k^*(t - t_1) \approx s_l(t_1) \approx s_l(t)$$

or that the resulting median point is being situated approximately on the slower varying component $s_l(t)$. When this is subtracted from the signal we get the fastest varying component of the signal $s_h(t) \rightarrow \text{IMF}$.

We proved that the fast varying component is being sifted out first in the EMD sifting process when the arbitrary slow varying component can be approximated by a piecewise straight line.

3.6.2 Piecewise Cubic Spline for Signal Envelope Construction and its Role in the EMD Sifting Scale Sequence for an Arbitrary Signal

The selection of a piecewise straight line approximation for the signal envelope boundaries $u(t)$, $d(t)$, except for very small time intervals δt used in the above conceptual proof, is unproductive for an EMD implementation. Instead the EMD uses *the piecewise cubic spline interpolation*. Surprisingly, the above examined linear approximations of the low varying component of a composite signal on small intervals of the argument prove useful, when considering the implications of piecewise cubic splines selection for envelope boundaries.

In the algorithm, we deal with discrete data and their envelopes being interpolated by EMD using a piecewise cubic spline function for adjacent extrema points of the same type, either maxima or minima. *To distinguish these extrema data points among all signal data $s(t)$ we will call them $x(t)$* . On an interval of two adjacent extrema of the same type $[x_i, x_j]$, the function $x(t)$ can be expressed by a cubic polynomial of variable t belonging to interval $[0, 1]$

$$x(t) = a + b*t + c*t^2 + d*t^3 \quad (10)$$

For small values of t (in the neighborhood of $t=0$) the second and third terms in $x(t)$ become very small and can be ignored. In general, $x(t)$ can also be expressed as a sum of two terms, yielding

$$x(t) = (a + bt) + (c*t^2 + d*t^3)$$

While the upper and lower envelopes represented by a piecewise cubic spline connecting maxima and minima extrema points on an interval $[t_1=t_i, (t_1 + \Delta t)=t_j]$ can be conceptually viewed as a line segment $(a + bt)$ or slow varying signal component modulated with a fast varying component $(c*t^2 + d*t^3)$. The same considerations hold for the extrema envelope lower boundary approximation by piecewise cubic splines.

When the median is then viewed as a cubic spline polynomial, as it actually is a difference of two envelope polynomials, it is also obvious that all three approximation lines - the upper envelope, lower envelope and the median line - are essentially parallel, with the median line following *the slower varying signal component trend*.

For a fast varying component the interval $[t_1=t_i, (t_1 + \delta t)=t_j]$ between two adjacent arguments is very small and the linear part of the piecewise cubic spline is the envelopes' and their median' approximation on the entire interval and the proof for the general case signal is similar to one provided above in Section 3.6.1.

However, two adjacent extrema points, as indicated by their two indices, of the same type x_i , $x_{i+1=j}$ may be widely separated among all signal data points and this case requires further considerations as follows. The coefficients $\{a, b, c, d\}$ for the piece of a cubic spline passing through point pairs x_i and x_{i+1} for parameter arguments $t_i=0$ and $t_{i+1}=1$ and $0 \leq i \leq N-1$ can be computationally determined from the four conditions for these two obvious end points and two more controversially selected slopes of $x(t)$ at interval ends, namely k_i and k_{i+1} :

$$\begin{aligned} x_i &= a && \text{(at given argument } t=0 \text{ and signal data } x_i) \\ x_{i+1} &= a + b + c + d && \text{(at given argument } t=1 \text{ and signal data } x_{i+1}) \\ k_i &= b && \text{(the first derivative of } x(t) \text{ } b + 2ct + 3dt^2 \text{ at } t=0) \\ k_{i+1} &= b + 2c + 3d && \text{(the first derivative of } x(t) \text{ } b + 2ct + 3dt^2 \text{ at } t=1) \end{aligned}$$

The solution for the cubic spline polynomial coefficients $\{a, b, c, d\}$ in terms of the two known extrema data point magnitudes x_i and x_{i+1} , and EMD process selected finite slopes k_i , k_{i+1} , are:

$$\begin{aligned} a &= x_i \\ b &= k_i \\ c &= 3(x_{i+1} - x_i) - 2k_i - k_{i+1} \\ d &= 2(x_i - x_{i+1}) + k_i + k_{i+1} \end{aligned} \quad (11)$$

For example, if we consider the two slopes being zero or $k_i = k_{i+1} = 0$ (as they should be for $s(t)$ at any extrema point argument), then the solution determines the following polynomial

$$x(t) = x_i + 3(x_{i+1} - x_i)t^2 + 2(x_i - x_{i+1})t^3 \quad (12)$$

or

$$x(t) = x_i + (x_{i+1} - x_i)t^2 * (3 - 2t) \quad (13)$$

If $x_i \approx x_{i+1}$, then the cubic spline polynomial (13) becomes a straight line $x(t) = \alpha x_i$. If the envelope becomes very symmetric and extrema of the same kind are close in value, then the cubic spline becomes very close to *straight line segments* $x(t) \approx x_i$ and this is resulting in fast varying component adjacent extrema (pair of a consecutive maxima and minima) to be selected as first invariants.

In the general case there are N points and $N-1$ piecewise splines. If we denote a piecewise spline that begins at point x_i as $Y_i(t)$, we have $N-1$ polynomials and need to determine $\{b_i, c_i, d_i, k_i\}$ for $0 \leq i \leq N-2$ or $4(N-1)$ unknowns. For this we will need $4(N-1)$ equations. System of equations (11) is not valid for the last polynomial and we need more conditions imposed on the polynomials. These are obtained by requiring the polynomial's continuity at joint points, the continuity of their first and second derivatives, as well as second derivatives being zero at the two data end points. This completes the specifications of the full system of linear equations for $\{a, b, c, d\}$ for $N-1$ cubic splines which has been proven to have a unique solution [10]. It is sufficient for our task to just know that this system has a solution and that the polynomials present a "natural" curve smoothly connecting the control data points. It is not obvious but the median $M(t)$ of two piecewise cubic splines $u(t)$ and $d(t)$ as defined in Section 2 equation (2) is also a piecewise cubic spline. Indeed, if we set

$$\begin{aligned}u(t) &= a_1 + b_1t + c_1t^2 + d_1t^3 \\d(t) &= a_2 + b_2t + c_2t^2 + d_2t^3\end{aligned}$$

then

$$M(t) = u(t) - d(t) = (a_1 + a_2)/2 + ((b_1 + b_2)/2)t + ((c_1 + c_2)/2)t^2 + ((d_1 + d_2)/2)t^3$$

and function $M(t)$ is obviously contiguous and differentiable and its first derivative is also continuous and differentiable with the resulting second derivative being contiguous. In other words, *function $M(t)$ is also a smooth function*. In the proximity of an extrema point $M(t)$ is close to the line of which the other end is the next opposite type extrema, making this pair an invariant for the following sifting steps. In the beginning of the EMD sifting process all the extremas obviously belong to the fastest varying component. Smooth connection of them by a piecewise cubic spline *results in a piecewise cubic spline median which is also smooth and as a consequence the fast varying adjacent extrema pairs are near symmetric or quickly molded (being made) into being symmetric and thus become first invariant pairs in the sifting process*.

This results in the fastest varying component of an arbitrary signal $s(t)$ to be sifted out first by the EMD sifting process.

Finally, what is important is to re-visit the concept of piecewise cubic spline smoothness by evaluating the polynomials' global maximum G on $[x_i, x_{i+1}]$

$$G = \max(s\{x_i, x_{i+1}, x(t_1^0), x(t_2^0)\}) \quad (14)$$

where t_1^0, t_2^0 are the two roots of the cubic spline polynomial (10) first derivative

$$b + 2ct + 3dt^2 = 0 \quad (15)$$

It is important that the piecewise cubic splines $u(t)$, $d(t)$ control points are the data extrema points $\{x_i\}$. This ensures that G is very close to $\max(s_1, s_2, \dots, s_N) = \max\{x_i\}$.

It is also, for example, important to work with near-zero slopes for which G is close to $\max\{x_i, x_{i+1}\}$, assuring near-linear behavior of the cubic spline between two sparse extrema points of the same type. For example, if

$$|k_i - k_{i+1}| = \varepsilon \text{ and } |x_{i+1} - x_i| = \varepsilon, \text{ where } \varepsilon \text{ is small}$$

Then c and d are also small and

$$x(t) = a + bt$$

G in (14) can also be evaluated using *the law of the mean from Calculus* [6], [11]. The first derivative of the cubic spline (15) is a *parabola* varying from b_i to $2c_i + 3d_i$ (for 0 and 1 end parameter values at $t=t_i$). The second derivative of the cubic spline is $2c + 6dt$ and is a straight line while it can only grow or only decrease on $[x_i, x_{i+1}]$ and it is definitely differentiable on the open interval (x_i, x_{i+1}) . The law of the mean states:

Let x be a function of t which is continuous at each point of the closed interval $A \leq t \leq B$, and let it have a derivative at each point of the open interval $A < t < B$. Then there is a point $t = \theta$ in the open interval ($A < t < B$) such that

$$x(t_{i+1}) - x(t_i) = (t_{i+1} - t_i) * x'(\theta)$$

The point θ is such that the derivative at θ gives a slope equal to the slope of the straight line connecting $x(t_{i+1})$ and $x(t_i)$ and which is the straight line around which the fastest varying component is locally weaving around on interval $[x_i, x_{i+1}]$. That the locally fastest varying component is being sifted out first in such a case has been already proven above.

This concludes the initial proof of Hypothesis 1 that for an arbitrary signal the fastest scale component is sifted out first by the EMD sifting process, and as a consequence, the remaining sifting becomes computationally less difficult. Other issues, such as the theory of local symmetry and IMF structure; the theory of the EMD rapid convergence; HHT signal synthesis; and ways of breaking up the signal for faster EMD processing, are left for a future paper in which we will prove the following Hypotheses 2 –3:

Hypothesis 2: *The EMD sifting process preserves an intermediate locally symmetric zero-crossing pair of extrema points with interleaved regions of diminishing amplitudes yielding an IMF with a definitive structure.*

Hypothesis 3: *The EMD sifting process' rapid convergence is of order $O(2^k)$. This is a consequence of the EMD envelope control points definition as sets of extremas of the same type, its interpolation by piecewise cubic spline whose control points are the data extremas, and envelope's median construction as an arithmetic median- the envelopes' sum divided by 2.*

Conclusions

We have reported the initial theoretical proof of why the fastest changing component of a composite signal is being sifted out first in the Empirical Mode Decomposition sifting process as implemented in the HHT-DPS. We have also provided the two hypotheses for the theoretical explanation of why the EMD algorithm converges and converges rapidly while using cubic splines for signal envelope interpolation. As part of this research we developed a few practical techniques for cutting a large input data set into smaller files, which facilitates faster HHT-DPS processing of large sound files. We have also developed in parallel with this investigation a few related to this research applications. For example, we have developed one of the first techniques for signal synthesis using the Hilbert Spectrum, by “painting” a recognizable 2-D pattern in Hilbert Spectrum image and then tracing back (direct inversion) the painted subset pixels to their origin in the IMFs, and deleting them from the IMFs. This then allows reconstructing the portion of the input signal (its synthesis) from the modified IMFs that depicts the extracted feature being successfully removed. These developments will be presented in follow up papers. Future work includes research on the IMFs basis orthogonality, the affects of normalization on the Hilbert Transform and resulting instantaneous frequency, as well as further research in the 2-D signal processing domain and handling of intermittency.

Acknowledgements

This work was performed by the authors as a team of NASA GSFC Applied Engineering and Technology Directorate (AETD) researchers and was funded by the AETD 2005 Research and Technology Development of Core Capabilities Grant (Advanced Communications R&TD 2005), as well as by the NASA Technology Transfer Accelerated Partnerships (TTAP) program. The assistance of Michael A. Johnson/AETD Code 560 Chief Technologist and Nona Cheeks/TTAP Code 504 Chief in sponsoring and encouraging this work is greatly appreciated and acknowledged.

References

- [1] The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analyses by Norden E. Huang et al. Proc. R Soc. London. (1998) 454, 903-995
- [2] A new view of nonlinear water waves: The Hilbert Spectrum by Norden E. Huang at al Annual. Rev. Fluid Mech. 1999, 31 417-457
- [3] Norden E. Huang Patent #6,311,130 B1
Patent Date: Oct. 30, 2001
Patent Name: Computer Implemented Empirical Mode Decomposition Method, Apparatus, and Article of Manufacture for Two-Dimensional Signals
- [4] Semion Kizhner, Thomas P. Flatley, Dr. Norden E. Huang, Karin Blank, Evette Conwell "On the Hilbert-Huang Transform Data Processing System Development Phase 1", 2003 MAPLD Conference Washington D.C., September 9-11, 2003
- [5] Semion Kizhner, Thomas P. Flatley, Dr. Norden E. Huang, Karin Blank, Evette Conwell "On the Hilbert-Huang Transform Data Processing System Development", 2004 IEEE Aerospace Conference Proceedings, Big Sky Montana, March 6-13, 2004
- [6] Angus E. Taylor, Advanced Calculus, Blaisdell Publishing Company, 1955, p.483
- [7] YA. Z. Tsypkin, "Sampling Systems Theory and its Applications Volume 1", The Macmillan Company, New York 1964
- [8] Robert W. Ramirez, "The FFT Fundamentals and Concepts", Prentice-Hall, 1985
- [9] Arthur Berry, "A Short History of Astronomy", Dover Publication 1960, par. 299
- [10] Teukolsky, Vettering, Flannery "Numerical Recipes in C The art of Scientific Computing 2d Edition", Cambridge University Press 1999
- [11] Beyer, "CRC Standard Mathematical Tables 26th Edition", CRC Press, Inc.