The analysis of the Empirical Mode Decomposition Method

Vesselin Vatchev, USC

November 20, 2002

The Empirical Mode Decomposition Method

For a given real signal $s(t)$ we look for a decomposition into simpler signals(modes)

$$s(t) = \sum_{j=1}^{M} a_j(t) \cos \theta_j(t),$$

where $a_j(t)$ is the amplitude and $\theta_j(t)$ is the phase of the j-th component. Each of the components has to have **physical** and **mathematical** meaning. Let $s(t)$ be "mono component" signal, i.e. we can find representation of the form

$$s(t) = a(t) \cos \theta(t)$$

that is both physically($\theta'(t) \geq 0$) and mathematically meaningful. There are infinitely many ways to construct such representations but it is often advantageous to write the signal in complex form

$$S(t) = s(t) + is_i(t) = A(t) \exp i\phi(t)$$

and to take the actual signal to be the real part of the complex signal. The imaginary part $s_i(t)$ of $S(t)$ has to be chosen to achieve a sensible physical and mathematical description. If we can fix the imaginary parts we can then unambiguously define the amplitude and the phase by

$$A(t) = \sqrt{s^2 + s_i^2} \ ; \quad \phi(t) = \text{arctan}\,\frac{s}{s_i}.$$

How to define the imaginary part?

- "quadrature method"

- analytic signal method - $s_i(t)$ is the Hilbert transform of $s(t)$.

"instantaneous frequency" of $s(t)$ is $\theta'(t)$

"instantaneous bandwidth" is $|\frac{a'(t)}{a(t)}|$.

"narrow band" instead of "mono-component".

The most popular example is

$$s(t) = A\cos\theta(t).$$

Some "paradoxes" exist. To avoid them, Huang, *et al* (N.E. Huang, Z. Shen, and S.R. Long, A new view of nonlinear water waves: the Hilbert spectrum, Annu. Rev. Fluid Mech. **31** (1999), 417–457) have developed a method, termed the "Hilbert view", for studying nonstationary and nonlinear data in nonlinear mechanics.

Main tools

- Empirical Mode Decomposition Method (EMD)

- Intrinsic Mode Functions (IMF's)

- Hilbert Transform.

Main idea

First identify the appropriate time scales that will reveal the physical characteristics of the studied system and then to decompose the function into modes intrinsic to the function. The EMD method was motivated "from the simple assumption that any data consists of different simple intrinsic mode oscillations".

- the time between successive zero-crossings;
- the time between successive extrema;
- the time between successive curvature extrema.

Intrinsic Mode Function (IMF)

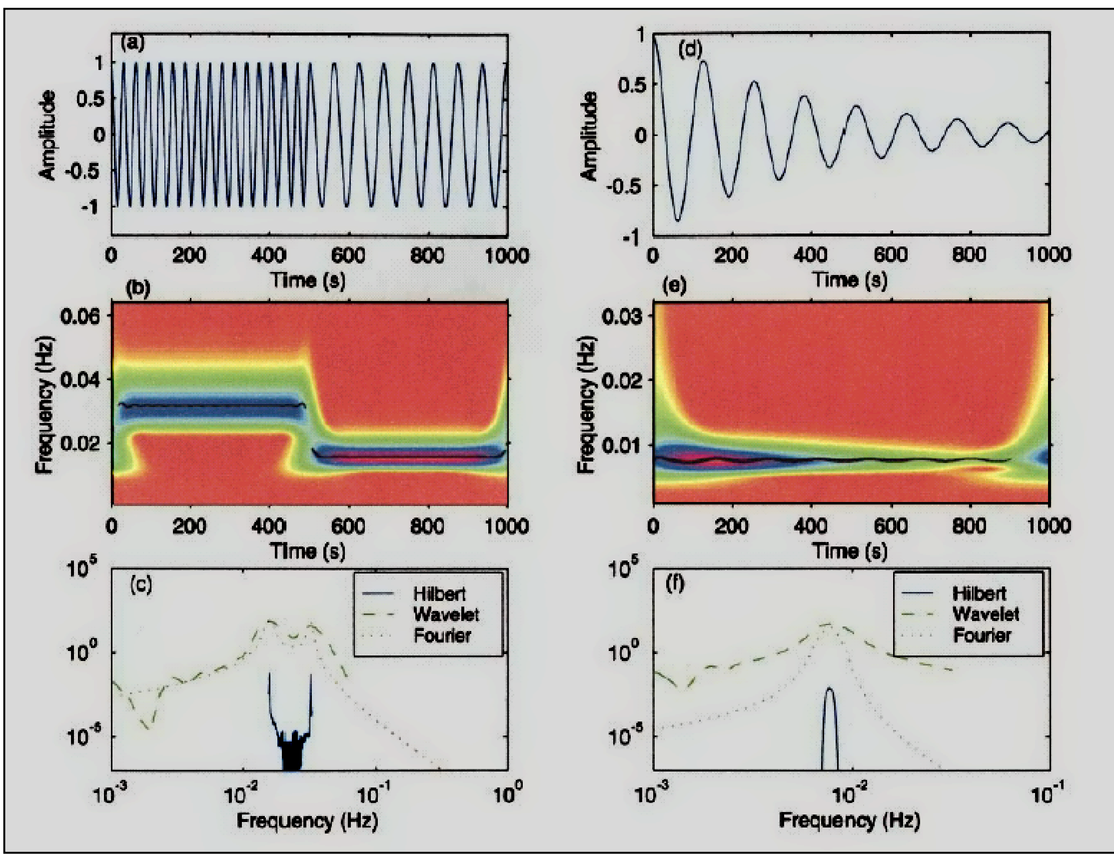
(a) The number of local extrema of ψ and the number of its zero-crossings must either be equal or differ at most by one.

(b) At any time t, the mean value of the "upper envelope" (determined by the local maxima) and the "lower envelope" (determined by the local minima) is zero.

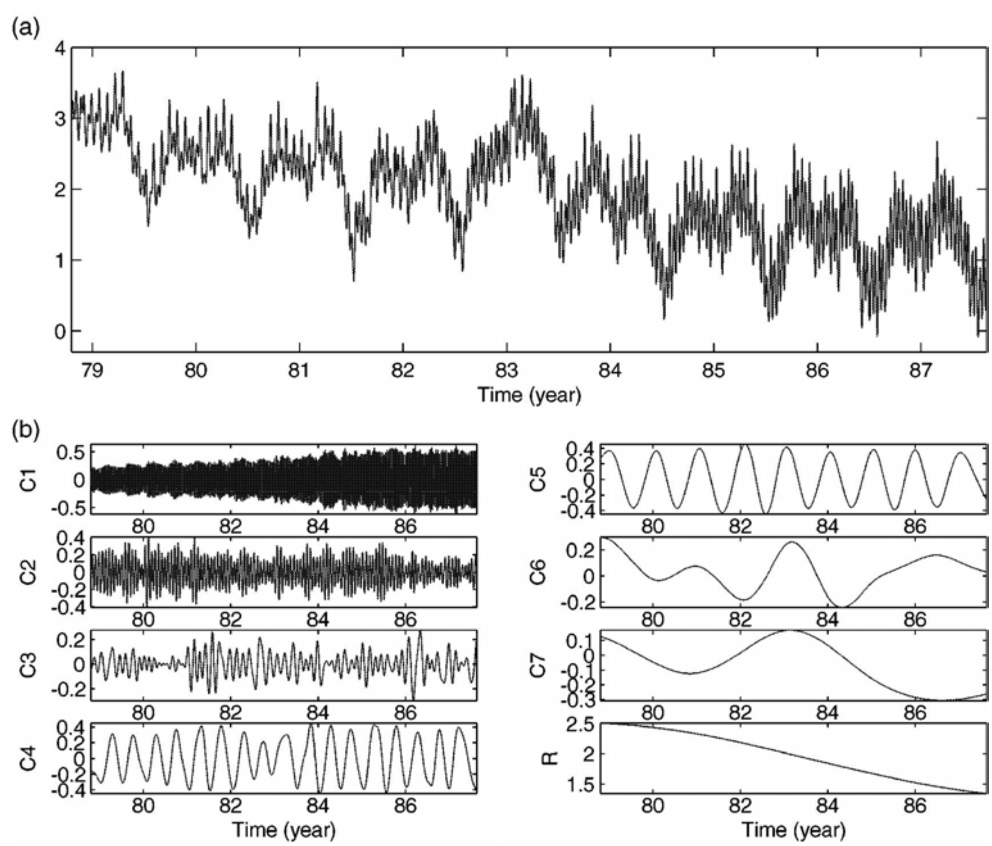
The "instantaneous frequency" obtained by applying the Hilbert transform to functions of the above type is very well localized in the time-frequency domain and reveals important characteristics of the signal. Figure 1 is reproduced from Huang *et al* and compares the Hilbert, Wavelet and Fourier spectra for a frequency and amplitude modulated signals that are IMFs. Figure 2 illustrates the EMD method for a physical signal of length of day.

Comparison of Spectra

Hilbert, Wavelet, Fourier



EMF Decompositon for Length of Day Data (Huang, et al)



The EMD Method to decompose f as a linear combination of IMF's ψ_n

Set $n := 0$ and $f_0 := f$.

Step 1. Set $h_0 := f_n$ and $k := 0$.

Step 2. *Construct the Upper Envelope for* h_k
Identify all the local extrema, then fit all the local maxima by a cubic spline interpolant for use as the upper envelope $U(t)$.

Step 3. *Construct the Lower Envelope for* h_k
Proceed in a similar way as Step 2, but replacing the local maxima as knots, by the local minima.

Step 4. *Sift*
Ideally, the functions U and L should envelop the data between them, i.e., $L(t) \leq$

$h_k(t) \leq U(t)$, for all t. Their mean is denoted by $m_k(t)$, and the k-th component is defined as

$$h_{k+1} := h_k - m_k.$$

If h_{k+1} is not an IMF,

then increment k, return to Step 2 and repeat the procedure (i.e., with h_{k+1} in place of h_k).

else define the IMF ψ_n as h_{k+1} and the residual f_{n+1} as $f_n - \psi_n$. If a convergence criteria is not met, increment n and return to Step 1. Convergence criteria typically consist of testing if the residual is either smaller than a predetermined value or is a monotonic function)

In view of the possible deficiency of the envelopes and to make the algorithm more efficient Huang *et al* suggest that the stopping criterion on the inner loop be changed from the condition that the “resulting function to be an IMF” to the single condition that “the number of extrema equals to zero-crossings” along with visual assessment of the iterates.

$$f(t) = \sum_{n=1}^{N} \psi_n + r_{N+1}$$

where $r_{N+1} = f_{N+1}$ is the residual.

$$f(t) = \Re \left(\sum_{n=1}^{N} A_n(t) \exp(i \int \omega_n(t)\, dt) \right).$$

The residue r_N is left on purpose, for it is either a monotonic function or a constant.” Here ω_n is the instantaneous frequency defined as $\frac{d\theta_n}{dt}$ where $\theta_n := \arctan(H\psi_n/\psi_n)$

Analysis and comments

The *first step* is to choose a time scale which is inherent in the function $f(t)$ and has physical meaning.

- the set of inflection points;
- the set of zeros of the function $(f(t) - \cos kt)$.

The *second step* is to extract some special (with respect to the already chosen time scale) functions, which in the EMD case are called IMF's. The definition of an IMF, although somewhat vague, has two parts (a) the number of the extrema equals the number of the zeros and (b) the upper and lower envelopes should have the same absolute value.

If we restrict to the larger class of functions which just satisfy condition (a), then we are able to provide a characterization in terms of solutions of self-adjoint ODE's.

Properties of the solutions of a self-adjoint ODE's.

An ODE is called self-adjoint if can be written in the form

$$\frac{d}{dx}\left(P\frac{du}{dx}\right) + Qu = 0, \qquad (1)$$

for $x \in (a, b)$ (a and b finite or infinite), where $P > 0$ and Q is continuous. These equation arise in vibration problems of continuum mechanics. The solutions are harmonic standing waves and it is commonly assumed that any wave motion can be resolved into simple harmonic waves (which can be proved mathematically).

I.(Max and Min) If $Q > 0$, then any solution of (1) has exactly one maximum or minimum between successive zeros. (This is the important condition for IMF's.)

II.(The Prüfer substitution) Let $Pu' = r(x)\cos\theta(x)$, $u(x) = r(x)\sin\theta(x)$, then (1) is equivalent to a linear system of first order ODE's:

$$\begin{aligned}\frac{d\theta}{dx} &= Q(x)\sin^2\theta + \frac{1}{P(x)}\cos^2(\theta)\\ \frac{dr}{dx} &= \frac{1}{2}\left(\frac{1}{P(x)} - Q(x)\right) r\sin 2\theta.\end{aligned}$$

If $Q(x)$ is positive, then the instantaneous frequency defined through the Prüfer substitution is always a positive number. If we want to pose some conditions on the envelopes (on the amplitude r in the above substitution), then they will affect only the second equation.

Relation between the self-adjoint ODEs and IMFs

Theorem. *Let* f *be a given* $C^2[a,b]$, $(a,b \in R^\infty)$ *and all of the zeros of* f *and* f' *be simple zeros. Then the following three conditions are equivalent:*

(i) *The number of the zeros of* f *and the number of the extrema of* f *on* $[a,b]$ *differ at most by* 1;

(ii) *There exist positive, continuously differentiable functions* P *and* Q *s.t.* f *is a solution of the self- adjoint ODE*

$$(P(t)f'(t))' + Q(t)f(t) = 0;$$

(iii) *There exists a* $C^2[a,b]$ *function* g *s.t.*

$$f(t) = -\frac{1}{Q(t)}g'(t), \quad g(t) = P(t)f'(t)$$

for some positive, continuously differentiable functions P *and* Q.

Idea of the Proof:
(ii) $->$ *(i)* Follows from the properties of the self-adjoint ODE's.

(i) $->$ *(ii)* # of zeros = # of extrema = M. Then consider the interpolation table

$$T = \left(\{t_j, |f(t_j)|, 0\}_{j=1}^{M}, \{z_j, |a_j|, 0\}_{j=1}^{M} \right),$$

t_j are the extremal points,
z_j are the zeros ($t_1 < z_1 < t_2 < z_2 < \ldots$),
and a_j are arbitrary positive numbers.

Hermite interpolant $E(f,t)_a$ for T s.t.

(a) The only extrema of $E(f,t)_a$ are at the points t_j, z_j for $j = 1, 2, \ldots, M$;

(b) The only points in common of the functions $E(f,t)_a$ and $f(t)$ are t_j, $j = 1, 2, \ldots, M$.

In general, any piece-wise function $E(f,t)_a$ made by pieces $\phi_j(t)$ that satisfy the conditions

$$\phi(y_1) = v_1,\ \phi(y_2) = v_2$$
$$\phi'(y_1) = \phi'(y_2) = 0,\ \ |\phi'(t)| > 0 \ \text{ for } \ t \in (y_1, y_2)$$

will work. (Meyer's wavelets.)

By choosing $a_j (j = 1, \ldots, M)$ we can insure that $E(f,t)_a$ satisfies (b) as well.

We have

$$g(t) := \frac{f(t)}{E(f,t)_a} = \sin \theta(t)$$

for some function $\theta(t)$ s.t. $\theta(t_j) = \frac{2k_j+1}{2}\pi$ and $\theta(z_j) = l_j \pi$ for some integers k_j and l_j.

Moreover we can adjust the a_j's so that $\theta(t)$ is C^1 function and $\theta'(t) > 0$ on the whole interval (a, b).

Let $r(t) := E(f,t)_a$ be an envelope for f with properties (a)-(b), and $\theta'(t) > 0$. Then the function

$$h(t) := r(t)\cos\theta(t)$$

has zeros at the points $t_j,\ j = 1, 2, \ldots, M$, and no other zeros on $[a, b]$.

The only zeros of $h'(t)$ on $[a, b]$ are the points $z_j,\ j = 1, 2, \ldots, M$.

By observing the changes of the signs of the functions f, f', h, and h', we can prove that

$$P(t) := \frac{h(t)}{f'(t)}, \quad Q(t) := -\frac{h'(t)}{f(t)}$$

are well defined, strictly positive($f(t_1) > 0$) $C^1[a, b]$ functions.

Comparing the relations $h'(t) = (P(t)f'(t))'$ and $h'(t) = -Q(t)f(t)$ we get

$$(P(t)f'(t))' + Q(t)f(t) = 0.$$

(iii) $->$ *(ii)* from the previous case.
(ii) $->$ *(iii)* from the Prüfer substitution

$$Pf' = r(t)\cos\theta(t), \quad f = r(t)\sin\theta(t).$$

Remark 1. The function g from *(iii)* also satisfies a self-adjoint ODE, namely

$$(\frac{1}{Q(t)}g'(t))' + \frac{1}{P(t)}g(t) = 0.$$

Remark 2. The coefficients P and Q can be represented by the amplitude $r(t)$ and the phase $\theta(t)$ in the following way

$$\frac{1}{P} = \theta' + \frac{r'}{r}tg\theta, \quad Q = \theta' - \frac{r'}{r}ctg\theta.$$

Moreover if $f(t_1) > 0$, then

$$\frac{1}{P} \leq Q,$$

and we have equality iff $r'(t) = 0$ on $[a, b]$.

The functions $E(f,t)_a$ considered in the theorem are a class of envelopes for f through its extrema. The choice of particular envelope (amplitude) $r(t)$ depends on the interpolation method and the vector a.

If we introduce some optimal condition $L(\cdot)$ we could specify only one function from that class and call it the envelope for f through the extrema.

Example: Let

$$SL(E) := \min_a \left(\int |E''(f,t)_a|^2 \, dt \right)^{1/2}$$

be the optimal functional. Then the envelope that minimizes $SL(E)$ should be very close to the natural cubic spline trough the points $(t_j, |f(t_j)|)$, $j = 1, 2, \ldots, M$.

Observation. The practical realization of the EMD method of Huang *et al* uses cubic splines as upper ($U(t)$) and lower ($L(t)$) envelopes. The nodes of these two splines interlace and do not have points in common so $U(t) = L(t) = at^2 + bt + c$. Then we have complete characterization of the IMF's.

Theorem. *If f is an IMF (so the upper and lower spline envelopes have same absolute value), then the amplitude is a constant.*
Theorem. *The amplitude is a constant for the associated self-adjoint ODE if and only if $Q(t) = 1/P(t)$.*

However, constant amplitude signals are usually not regarded as an acceptable class of functions for IMF's. If we drop (b), we will have a reasonable (from practical point of view) definition but, at the next stage, unrecoverable mathematical ambiguity in determining the modulus and phase will be introduced. In any case, the definition of IMF must include condition (a).

Definition. *A twice differentiable function f is an IMF if it is a solution of the self-adjoint ODE of the type*

$$(Pf')' + Qf = 0,$$

for some $P(t) > 0$, $Q(t) > 0$ for $t \in [a, b]$

and . . .

To overcome the envelope restriction, we can either modify the construction of the envelopes or, instead of requiring $U(t) = L(t)$ for all t, we can require $|U(t) - L(t)| \leq \epsilon$, for some prescribed $\epsilon > 0$.

A condition for symmetry could be found by considering the bandwidth $|\frac{r'(t)}{r(t)}|$.

Current problems.

?1. The main problem is how the properties of the envelopes are related to the Hilbert transform of a given function.

?2. In many problems an amplitude and a phase are defined through the Prüfer substitution. The two first order linear equations define the instantaneous frequency and the instantaneous bandwidth in terms of the coefficients. What is the physical background of that approach and how it is related to the Hilbert transform?

?3. What is the exact criteria for envelope through the extrema? What are upper and lower envelopes?

?4. Substituting property (b) from Huang *et al* by some property that is not that restrictive? Does the Prüfer method provide an appropriate substitution?

?5. To continue studying the sifting step and to consider appropriate modifications (e.g. cubic spline through the inflection points, $f'' + Qf = 0$).

?6. To define a class for which the Greedy algorithms will produce meaningful representation and to find the order of convergence of the EMD method in appropriate approximations spaces.