A formal study of the nonlinearity and consistency of the Empirical Mode Decomposition

Nikolaos Tsakalozos, Konstantinos Drakakis*, Scott Rickard

UCD CASL, University College Dublin, Belfield, Dublin 4, Ireland

**Article info**

Article history:
Received 16 February 2011
Received in revised form
12 September 2011
Accepted 13 September 2011
Available online 19 September 2011

Keywords:
Empirical Mode Decomposition (EMD)
Intrinsic Mode Functions (IMFs)
Envelope
Consistency
Nonlinearity

**Abstract**

Towards developing a rigorous mathematical theory for Empirical Mode Decomposition (EMD), we provide an overview of the algorithm and introduce a corresponding operator, attempting a preliminary study. We prove that the EMD is nonlinear, we identify the major reason of its nonlinearity, and we introduce the related concept of consistency, which we show the EMD does not satisfy either.

© 2012 Published by Elsevier B.V.

1. Introduction

Smooth and periodic signals defined on the real line can be expanded into a Fourier series of sines and cosines whose frequencies are integral multiples of the inverse length of the period. This expansion has two main and often desirable features: (a) the terms of the expansion are orthogonal to each other and (b) the collection of functions over which the expansion takes place is chosen before, and without any reference to, the function being expanded. Thus, sines and cosines form a “dictionary” over which functions can be compressed without any loss of information: signals that seemed before to need an uncountable infinity of values in order to be fully specified can now be encoded through a countable infinity of (or even finitely many) real numbers, namely the coefficients of the trigonometric functions.

This representation, however, seems at times to be excessively rigid. Consider, for example, a pure sine function $y(t) = \sin(2\pi ft)$, whose Fourier expansion trivially contains one term, and consider further its slight perturbation $(1 + \epsilon_1(t))\sin(2\pi f(1 + \epsilon_2(t))t)$, where $\epsilon_{1,2}$ are functions taking “small” values but such that the perturbation is still periodic of the same period as $y$. As a consequence of the inflexibility of the trigonometric dictionary, this perturbation will most likely lead to an expansion containing infinitely many terms, despite being so similar to the pure sine $y$.

There are, moreover, numerous occasions (see, for example, [17] in the context of image processing) where the orthogonality of the terms of the expansion is of lesser or no practical interest. Indeed, a fixed and a priori chosen dictionary, despite being convenient and simple, often diffuses useful information about the signal over a large number of expansion terms, thus making it more difficult to notice. In dealing e.g. with non-stationarity, sparse and flexible expansions ensure that this information is captured in a small number of terms.

The Empirical Mode Decomposition (EMD) is a relatively recently proposed algorithm, formulated by Huang et al. [21] in 1998, specifically addressing these concerns. Given a function (not necessarily periodic), it produces a “Fourier-like” expansion over “trigonometric-like” functions, called...
Intrinsic Mode Functions (IMFs), constructed out of the function \( f \) itself; in fact, the class of IMFs contains all trigonometric functions but is significantly more general, and this extension of the set over which expansion terms are chosen can lead to sparser expansions. Moreover, the IMFs have indeed been shown to contain useful information about the function in a variety of contexts and occasions ([9,15,21,35]) and furthermore [5,28] for more “exotic” applications.

The downside of the EMD is its empirical nature: despite the efforts of numerous researchers [21,23,34,37], it remains an exclusively empirical algorithm without a solid mathematical foundation. Contrary to Fourier or wavelet transforms, then, where comparison of implementation performance against the rigorous definition is a non-issue, as flawless numerical implementations of the transforms have been available long ago, in the case of EMD it is practically impossible to tell which aspects of its performance are to be attributed to the concept or its implementation. For example, EMD seems to rely heavily on the concept of the envelope of a function, and its performance apparently varies drastically depending on the particular choice one makes for it [35]. Yet, it is not clear whether a particular envelope should be considered as a part of the formal EMD definition or a mere feature of each implementation.

In this work we take some steps towards a formal foundation and study of the EMD. We introduce an operator for it, and we make some remarks about its domain and range. We then offer a formal proof of the nonlinearity of the EMD, and we recast the definition of linearity of a transform into the implementation-verifyable equivalent definition of consistency: does the transform decompose a linear combination of components, each of which remains fixed under the action of the transform, back into the components themselves? Though, in view of its nonlinearity, EMD cannot, of course, satisfy this property, it might satisfy a weaker requirement: assuming a transform decomposes a signal into some components, does it decompose any linear combination of those particular components back into themselves?

A first discussion of the issues considered herein was presented by the same authors in ISSC 2009 [10]. Although there is significant overlap of the material covered in the first three sections of each paper, the conclusions are significantly different. In particular, [10] formulated the concepts and problems studied here and proposed a methodology to address them, hence it could be construed as a study plan for the current work.

2. Description of the EMD

Consider a continuous and smooth real-valued function \( f \) defined on a finite closed interval \( D \subset \mathbb{R}, \) with a finite number of local minima and maxima. By convention, the endpoints of \( D \) will not be considered amongst the local minima and maxima, but will be treated separately instead. The EMD produces some new functions \( f_i, \) \( i = 1, \ldots, N \) and \( r, \) where \( N \) is not known a priori but rather determined by the algorithm, such that

\[
f = \sum_{i=1}^{N} f_i + r.
\]

The functions \( f_i, i = 1, \ldots, N, \) oscillatory in nature, are the IMFs of \( f, \) while \( r \) is known as the trend of \( f, \) which does not perform a full oscillation over \( D, \) namely it has at most one extremum in it.

Mainstream versions of EMD rely crucially upon the extraction of the envelope of \( f. \) We therefore proceed to discuss this concept first, before offering any more details on the algorithm itself, the IMFs, and the trend.

2.1. Envelopes

**Definition 1.** The envelope of a function \( f \) defined as above consists of two functions, the upper envelope \( f_u \) and the lower envelope \( f_l. \) They are required to satisfy the following conditions:

* For any \( x \in D, \) \( f_u(x) \leq f(x) \leq f_l(x). \)
* \( f_u(x) = f(x) \) whenever \( x \) is a local maximum of \( f, \) and similarly \( f_l(x) = f(x) \) whenever \( x \) is a local minimum of \( f. \)
* Whenever \( x_1 < x_2 \) are two consecutive local maxima, \( f_u(x) \) must lie between \( f_u(x_1) \) and \( f_u(x_2) \) for all \( x \in [x_1, x_2]. \) Similarly, whenever \( x_1 < x_2 \) are two consecutive local minima, \( f_l(x) \) must lie between \( f_l(x_1) \) and \( f_l(x_2) \) for all \( x \in [x_1, x_2]. \)
* For any \( \lambda > 0, \) \( \lambda f_u = f_u \) and \( \lambda f_l = f_l. \)

All four conditions should appear intuitively justifiable to the reader, despite not having appeared in such an explicit form elsewhere in the literature. The first three conditions ensure that the envelope bounds the signal as tightly as possible, avoiding unreasonably high or low values (this necessity was also recognized in [25]). In practice, the first and third conditions are only (required to be) approximately satisfied, while the second condition suggests that the envelope can be constructed through an interpolation method relying only on the positions of the local minima and maxima of \( f, \) on the endpoints of \( D, \) and on the values of \( f \) at these points. In other words, letting \( U(f) \) and \( U(f) \) be the positions of the local maxima and minima of \( f, \) respectively, and \( D = [d_1, d_2], \) we may set:

$$f_u = \mathcal{L}_u(U(f), f(U(f)(d_1), f(d_1)), f(d_2))).$$

$$f_l = \mathcal{L}_l(U(f), f(U(f)), d_1, d_2, f(d_1), f(d_2))).$$

where \( \mathcal{L}_u \) and \( \mathcal{L}_l \) are two (possibly different) interpolation laws producing functions defined over the entire \( D \) out of the finite sets of points passed to them as arguments. Having determined the upper and lower envelope, the local mean of \( f \) is defined as \( \bar{f} = (f_u + f_l)/2. \)

The fourth condition simply states that the envelope of a signal must scale along with the signal itself. This may sound so obvious that it is not even worth mentioning, but let us note that, perhaps surprisingly, non-scalable envelope constructions have, in fact, been proposed [33]!

These conditions are clearly insufficient to specify the envelope of \( f \) uniquely and unambiguously. However, the

---

1 This section is a revision of Section II in [10].
research community seems to have reached a consensus that the preferred construction method is cubic spline interpolation [15,21,23,34,35].

2.2. IMFs

IMFs [21,37] can be construed as generalized sines, specifically as sines with “variable width” and “fluctuating frequency”. The following definition is inspired by the results of [41,44]:

**Definition 2** (Intrinsic Mode Function). A function \( f \) defined on \( D \) is an Intrinsic Mode Function (IMF) if and only if there exist functions \( m \) and \( \phi \), and reals \( a \leq b \) (either of them allowed to be \( \pm \infty \) but not both) such that:

\[
 f(x) = m(x)\cos(\phi(x)), \quad x \in D,
\]

where \( \supp(m) \subset [a,b] \) and \( \supp(\cos(\phi)) \subset \mathbb{R}[-b,-a] \), \( \ast \) denoting the Fourier transform. In particular, \( |m| \) is called the envelope and \( \phi \) the phase of the IMF.

**Remark 1.** IMFs are the only functions for which the envelope has been rigorously defined [34]. Using the Bedrosian equality, the Hilbert transform \( \mathcal{H} \) [21,41] of an IMF can be shown to be [41]:

\[
 \mathcal{H}(m \cos(\phi)) = m \mathcal{H}(\cos(\phi)) = m \sin(\phi),
\]

whence the analytic signal of \( f \) becomes \( f + \mathcal{H}(f) = m \exp(i\phi) \), and the upper/lower envelope \( f_u/f_l \) of \( f \) can be defined as the amplitude/negative amplitude of the analytic signal, respectively: \( f_u = m, f_l = -m \). It follows that the local mean of an IMF is 0. Nevertheless, though the spectral conditions given in **Definition 2** are sufficient for Bedrosian’s equality to apply, to the best of our understanding it is still not known whether they are necessary.

In a simpler, imprecise, yet intuitive way, IMFs are usually defined [21,34,37] as functions having the following two properties:

**Definition 3** (Alternative IMF definition).

- All local maxima are positive and all local minima are negative.
- The local mean is 0.

Is this equivalent to **Definition 2**? It is relatively simple to show that an IMF as per **Definition 2** satisfies both properties given in **Definition 3**. Conversely, though, despite the fact that a function \( f \) satisfying both conditions given in **Definition 3** must be expressible in the form \( f = m \cos(\phi) \) for two functions \( m \) and \( \phi \) (where \( \phi \) is unique given \( m \), but \( m \) is not uniquely chosen given \( f \)), it is still not clear why Bedrosian’s equality should be applicable to it, so that \( m \) can indeed be identified as the envelope. Considering all pairs \( (m,\phi) \) for which \( f = m \cos(\phi) \), does a pair satisfying the spectral conditions given in **Definition 2** always exist? Even if it does not, can such a pair be found satisfying Bedrosian’s equality? If so, is this pair unique? These are important questions to which, to the best of our knowledge, no definitive answer can be given at present.

A rigorous mathematical theory has been developed for functions satisfying the first condition in **Definition 3**, but not necessarily the second, called weak IMFs. Specifically, it was shown that a function is a weak IMF if and only if it satisfies self-adjoint ODE of the Sturm–Liouville type [37,40]. Since it is the second condition (about zero local mean) that involves the envelope, this approach suggests that the envelope may be a hindrance to the mathematical study of IMFs, and that mathematical progress may be possible provided envelopes are kept out of our considerations.

In practice, \( f \) is accepted as an IMF whenever it satisfies **Definition 2** in an approximate sense. For example, for a pre-specified \( \epsilon > 0 \), the (interpolation-based or otherwise practically/numerically computed) local mean is required to satisfy \( \| (f_u + f_l) / 2 \| < \epsilon \), where \( \| \cdot \| \) denotes some function norm (such as the absolute maximum \( \| \cdot \|_{\infty} \) or the mean square value \( \| \cdot \|_2 \)) [21,23,35].

2.3. The EMD

Having discussed the envelope and defined what an IMF is, we are now ready to describe the EMD algorithm. Choose \( \epsilon > 0 \) and initialize by setting \( f_{1,1} = f \) and \( n = m = 1 \); then, perform the steps below:

1. Compute the upper and lower envelopes \( f_{u,n,m} \) and \( f_{l,n,m} \) of \( f_{n,m} \) and the local mean \( l_{n,m} = (f_{u,n,m} + f_{l,n,m}) / 2 \).
2. If \( \| l_{n,m} \| < \epsilon \), set \( f_{n} = f_{n,m} \); otherwise, set \( f_{n,m+1} = f_{n,m} - l_{n,m} \), increase \( m \) by 1, and go to Step 1.
3. Set \( f_{n+1,1} = f_{n,1} - f_{n} \); if this has at most one local extremum, set \( f = f_{n+1,1} \) and stop; otherwise, increase \( n \) by 1, reset \( m \) to 1, and go to Step 1.

Two empirically observed facts about the algorithm guarantee that the algorithm terminates:

- For all valid \( n \), \( \lim_{m} f_{n,m} \) is an IMF.
- For all valid \( n \), \( f_{n+1,1} \) has strictly fewer local extrema compared to \( f_{n,1} \).

No rigorous proof of these facts seems to be available in the literature. Regarding the second, in [35] the authors claim that the number of local extrema in residual functions progressively decreases “by construction”, offering no proof or reference, though.

2.4. Some remarks on envelope choice

The envelope, needed for the extraction of the local mean, is an indispensable tool in the current mainstream formulation of the EMD algorithm, yet, at the same time, it seems to cause the most complications in all attempts of a formal definition. A simple, effective, and popular choice for the envelope construction is the cubic spline interpolation, which has been widely recognized to yield very good results in practice [21,23,34,35]. In contrast, many other envelopes tested were found to “over-decompose” the
function into a large number of IMFs, increasing the number of iterations required for a given accuracy and decreasing the amount of information each IMF carries [35].

The cubic spline is not, strictly speaking, an acceptable envelope. Although it matches the function values on the interpolation points, and its first and second derivatives are continuous, the values of either derivative do not necessarily match the values of the corresponding derivative of the function at the interpolation points. Consequently, the envelopes do not lie tangentially above and below the function graph, as they should, but usually cross it (see, for example, [24]), thus violating the first requirement of Definition 1. A careful reading of the literature reveals that some authors avoid stating this requirement explicitly, referring instead simply to the upper and lower envelope as interpolating functions [11,34,35], while others (most notably Huang himself) [21,24] demand it, but comment no further on the possible contradiction it might cause.

It is often argued (see e.g. [13]) that the EMD is a local technique. Assuming that \( f \) is defined on \( D \) and that \( \varepsilon \) is defined on an interval \( E \subset D \), there exists an interval \( \tilde{E} \) such that \( E \subset \tilde{E} \subset D \), and such that the IMFs of \( f \) and \( f + \varepsilon \) are identical over \( D \tilde{E} \). This property, however, cannot possibly hold in the strict sense independently of the envelope used. Using cubic spline interpolation, for example, local changes propagate throughout the entire envelope due to the continuity conditions at the endpoints of the interpolation intervals (namely the extrema). It may well be that the effect of this propagation decays with the number of iterations performed to obtain an IMF, but, once more, this requires further investigation and a formal proof.

The envelope’s behavior near the endpoints of the observation interval of the function, where inevitably some degree of extrapolation is involved in order to complete its construction, is another important aspect of locality. In the case of the cubic spline (and many other splines), the decisions made there can potentially affect the envelope in the interior of the interval, propagating through the spline smoothness conditions imposed at the local extrema: this issue was recognized already in [21]. An excellent account of the different splines and different endpoint conditions used with the EMD, along with many references, can be found in [26,27]. Mirroring the function across the left and the right boundary, at least up to the first (leftmost and rightmost, respectively) extremum present, has been reported to lead to good extrapolation results [35,45].

Envelope issues have inspired several very interesting and radically different approaches to “traditional” EMD. For example, in [26,27], the authors propose using rational functions as interpolants instead of polynomials (these were also found to fail the first condition of Definition 1, though). As another example, out of a given signal \( f \), a new “masking signal” \( s \) can be appropriately constructed using information from \( f \), in order to act as a “catalyst” in the EMD decomposition of \( f \). EMD is used to extract the first IMF \( z_+ \) of the signal \( f_+ = f + s \), and the first IMF \( z_- \) of the signal \( f_- = f - s \); then, the first IMF \( z \) of \( f \) is given by \( z = (z_+ + z_-)/2 \). The remaining IMFs are extracted similarly, by choosing an appropriate masking signal at each step [11]. Alternatively, the local mean can be defined without resorting to the construction of envelopes, through the interpolation of the inflection points of \( f \) [24]. Finally, even when envelopes are used, they do not have to be constructed by interpolating the extrema of \( f \). This leads to the problem of selecting the optimal (in some sense) set of interpolation points. One possibility to achieve this is stochastically, using genetic algorithms [24].

The huge amount of published work on envelopes indicates the importance the research community attaches to this concept. Understanding fully the role of the envelope, and selecting and incorporating the best envelope in the method will constitute a major step towards the formalization of EMD. This is, however, not an easy task, as it requires answering many difficult questions. What is the class of possible envelope functions and how can the optimal envelope be selected from within this class? Does the optimal envelope depend on the signal? Can localization and (asymptotic) enfolding be proved for a given envelope?

Finally, note that it is the local mean that is necessary for EMD, and that envelopes are only used as an intermediate tool to extract the local mean. Can the definition of the local mean be somehow based directly on the signal? This approach, which, if successful, will bypass all envelope-related issues, has been investigated through three alternative avenues. In the first, the local mean was defined through the interpolation of the inflection points of the signal [24]. In the second, it was defined through a fourth order partial differential equation (PDE) [6], but success was, in both cases, limited. Finally, a much more promising definition of the mean through a parabolic (heat-type) PDE was recently proposed for both 1D and 2D signals [7,8].

3. Formalization of the EMD

The EMD expands functions as a sum of IMFs (and a trend) and is, therefore, an operator, which we will henceforth denoted by \( \mathcal{E} \). Postponing temporarily the question of the exact form of \( \mathcal{E} \), which remains the ultimate goal of our investigation, we still have to address the issues of its domain and range. Regarding the domain \( \mathbb{D} \), to the best of our knowledge, there is no rigorous definition in the literature, while there is admittedly little point in trying to deduce the broadest possible domain of an operator whose exact form and properties we are not aware of, so we are guided in our choice mainly by the requirement that the domain be broad enough to cover the documented instances in the literature of (real) functions the EMD has been applied to. Thus, we consider the domain to consist of all real, continuous and piecewise smooth functions defined on an interval, possessing a finite number of local extrema. Trivially, this can be expanded to include periodic functions defined over \( \mathbb{R} \), since this is equivalent to considering a single period (possibly by imposing appropriate boundary conditions at the ends of the interval). Allowing infinitely many local extrema (except for periodic functions) adds an unnecessary complication, as do discontinuous functions, and to the best of our knowledge no
such cases have yet been considered in the literature. Furthermore, we restrict our attention to the simplest case of the real line, leaving at this time extensions e.g. to the complex domain [39], to higher dimensions [18,38], or bivariate, trivariate, and generally multivariate functions [36,30,31] outside the scope of this work.

In order to define the range, let us first define the set $\mathcal{M}'$ of all IMFs and trends; note that IMFs can be defined rigorously through Definition 2, avoiding any reference to local means and envelopes. This is the set over which EMD expands functions. For reasons of economy, however, we can shrink this set further: expansions over this set would use binary coefficients, in the sense that an element of $\mathcal{M}'$ is or is not present in the expansion. This, however, is not how expansions are usually considered: for example, when we expand periodic functions of period $2\pi$ in a Fourier series, we consider $\sin(x)$ to be an element of the range of the transform, while we do not consider e.g. $5\sin(x)$ to be a different element of the range, but rather the previous element with a different coefficient. Likewise, we would like to view $M$ and $\lambda M$, $\lambda \in \mathbb{R}$, $M \in \mathcal{M}'$, as the same element of the range multiplied by two different coefficients rather than two different elements of the range.

For this reason, we now divide $\mathcal{M}'$ into equivalence classes: for any $M \in \mathcal{M}'$, its equivalence class is $\langle \lambda M : \lambda \in \mathbb{R} \rangle$. We subsequently form the set $\mathcal{M}$ by selecting precisely one representative from each equivalence class, say the one whose absolute maximum equals 1. $\mathcal{M}$ is an uncountably infinite set, indexed by some index set $A$: $\mathcal{M} = \{M_a : a \in A\}$. The range $R$ of $\mathcal{E}$ consists of those vectors $c = (c(a) : a \in A)$ with real coordinates which act as IMF coefficients in the expansion of functions lying in the domain, and is thus a subset of the set $\mathcal{C}$ of all vectors indexed by $A$ with real coefficients:

$$\mathcal{E}(f) = c \Rightarrow f = \sum_{a \in A} c(a)M_a.$$  

(5)

Assuming it holds true that, after the extraction of an IMF, the remainder has fewer local extrema than the original function, we conclude that only finitely many coefficients of a vector $c$ in the range of $\mathcal{E}$ can be nonzero, which is a nice feature of this formalization. Furthermore, at most one nonzero coefficient in $c$ will correspond to a trend, as, by the very description of the EMD algorithm, the trend is defined to be the non-oscillating remainder after all IMFs have been extracted. Hence, we refrain from using equivalence in (5), as there may be many coefficient vectors $c$ reconstructing $f$, only one of which is returned by the algorithm and thus lies in $\mathcal{R}$.

As an example of this formalization, the question of EMD linearity can be stated as follows: is it true that for any $f_1, f_2 \in \mathcal{D}$ and any $s_1, s_2 \in \mathbb{R}$, $\mathcal{E}(s_1f_1 + s_2f_2) = s_1\mathcal{E}(f_1) + s_2\mathcal{E}(f_2)$?

4. EMD nonlinearity and consistency

It is very frequently quoted in the EMD literature, in various but clearly equivalent forms, that “the EMD is a nonlinear process suitable for the decomposition of nonlinear and non-stationary data” or “signals” [3,5,13,20–22,24,27,35]. Though this statement is apparently satisfactory to a large part (possibly the majority) of the research community, and certainly adequate for applications, a slight misunderstanding could arise because of the use of the term “nonlinear signal”, which not only does not appear to have a standard definition in the literature of signals and systems, where linearity is a well defined property of systems rather than signals, but, what is more, carries an obvious alternative meaning.

We remind the reader that a function $f : \mathbb{R} \to \mathbb{R}$ is linear iff $\forall x,y \in \mathbb{R}, f(x+y) = f(x) + f(y)$ (this equation is often known as Cauchy's functional equation), and $f(xy) = f(x)f(y)$, and that, assuming further that $f$ is continuous, the only possible form of $f$ is $f(x) = Cx$ for some $C \in \mathbb{R}$ [1]. Upon encountering the term “nonlinear signals” then, one might jump to the conclusion that it describes any signal not of the aforementioned form. In the context of EMD, however, it appears that the term is used to describe data sets (time series) resulting as the output of a nonlinear process (system) [21,22], which is indeed a well defined concept.

Given, however, a data set (a signal), can we determine with certainty whether it is the output of a linear or a nonlinear system? This problem has been investigated in the past, and several tests have been proposed [2,20,22,29], which unfortunately yielded only necessary conditions. A closely related issue is the determinism/predictability of a nonlinear system? This problem has been investigated in the past, and several tests have been proposed [2,20,22,29], hence both transforms are time-scale transforms. The crucial difference between them, however, is that the (allegedly) nonlinear EMD produces one IMF per scale [19,22] (at least roughly so, as different IMFs may have slightly overlapping spectra [42]), which therefore contains all information regarding this scale, while the (linear) wavelet transform produces in general many terms living in a given scale, hence any information pertaining to a given scale is potentially diffused over a large number of terms and lost.

4.1. A formal proof of EMD nonlinearity

Theorem 1. Under the assumptions made above, EMD is nonlinear.

Proof. Consider the signal $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \cos(x) + \epsilon \cos(3x)$. Note that $f$ has two components, both of which are fixed under the action of the EMD, as they obviously satisfy Definition 3. We will show, however, that $f$ satisfies Definition 3 as well, hence $f$ itself is an IMF. We obtain:

$$-f'(x) = \sin(x) + 3\epsilon \sin(3x) = \sin(x) + 3\epsilon[3\sin(x)\cos^2(x) - \sin^2(x)].$$

$$= \sin(x)[1 + 3\epsilon(3 - 4\sin^2(x))].$$

(6)

But $1 + 3\epsilon(3 - 4\sin^2(x)) \geq 1 - 3\epsilon > 0$, assuming $\epsilon < 1/3$, so that $f'(x) = 0 \Leftrightarrow \sin(x) = 0 \Leftrightarrow x = k\pi$, $k \in \mathbb{Z}$. This implies that
\( f(2k\pi) = 1 + \epsilon = -f(2k\pi + \pi), \) so not only are the values of \( f \) at all maxima equal and positive, and at all minima equal and negative, but the two values are additive inverses. Consequently, combining the first three properties of Definition 1, we find that the constant functions \( 1 + \epsilon \) and \(- (1 + \epsilon) \) form the upper and lower envelopes of \( f \), respectively, so the local mean of \( f \) is 0. Thus, both conditions of Definition 3 are satisfied. This completes the proof. \( \square \)

Note that we have just demonstrated that the EMD “cannot separate tones”, but this is not meant to be understood as a weakness. Simply put, EMD should not be used in place of the Fourier transform in order to do what the Fourier transform does best, namely produce an expansion of pure tones. Instead, we should use the EMD when we are willing to accept that signals consisting of multiple pure tones have a better physical meaning than their constituent tones. In other words, paraphrasing [19], the choice of the right transform depends on our own anticipation of the meaning of the resulting expansion.

### 4.2. Consistent expansion into IMFs

It is our feeling that the question of linearity of an implementation-based transform is better addressed through an implementation-oriented recasting of the traditional definition of linearity. We thus arrive at the concept of consistency, which not only can be proved to be equivalent to linearity for linear transforms, but also allows the definition of a range of weaker linear-like behaviors for nonlinear transforms. Indeed, in any expansion rule, we recognize two fundamental types of consistency: input–output and output–output consistency. We proceed to formulate their definitions over discrete expansions for simplicity; regarding expansions over a continuum, just substitute summation by integration.

**Definition 4 (Input–output (IO-) consistent expansions).** Let \( T \) be an expansion rule, and let \( f \) be the sum of a collection of linearly independent functions \( \{f_i, i \in I\} \) belonging in its range, each of which is expanded by \( T \) into itself; \( T \) will be called IO-consistent if and only if each such \( f \) is expanded back into \( \{f_i, i \in I\} \).

Trivially, assuming \( T \) is linear, it is also IO-consistent. Conversely, if \( T \) is IO-consistent, and has, additionally, the property that \( T(\lambda h) = \lambda T(h) \) for any \( \lambda \in \mathbb{R} \) and any \( h \in \mathcal{D} \), then it is also linear.

As the EMD is not linear, it cannot be IO-consistent. Consequently, we will now subject the EMD to a further, weaker consistency test, conditioning consistency on the expanded signal. More precisely, after decomposing a signal into its constituent IMFs, we add together a subset of these IMFs, and we let EMD decompose this sum; will we recover the IMFs we added? Intuitively, we would like this to hold: expansions should not be affected by terms not included in the expansion, in the sense that “what we do not know should not hurt us”!

**Definition 5 (Output–output (OO-) consistent expansions).** Let \( T \) be an expansion rule which expands \( f \) into a set of functions \( \{f_i, i \in I \subset \mathbb{N}\} \), in the sense that \( f = \sum_{i \in I} f_i \); \( T \) will be called OO-consistent if and only if, for any such \( f \) and any \( j \in I \), \( T \) expands \( \sum_{i < j} f_i \) back into the functions \( f_i, i \in J \).

Note that an IO-consistent(linear) expansion is necessarily OO-consistent, but it is not obvious that the converse is true: therefore, OO-consistency is not stronger than IO-consistency(linearity). We study OO-consistency through a numerical simulation. To facilitate the reproducibility of our results, we use the EMD Matlab implementation by Flandrin et al., which is freely available on the web [12]. An important feature of this implementation are the iteration stopping criteria it contains, namely the parameters \( \text{stop and max iterations;} \) we use the default stopping criteria of the code, as they were chosen to be, in some sense, optimal [35]. In general, in order to process real world signals with EMD, an enhanced version of the algorithm should be used, known as Ensemble EMD (EEMD) [43], which is more robust to noise, which occurs naturally in the quantitative description of all physical processes. In our experiments below, however, we will be using artificial smooth signals, so this improved version of EMD will not be needed.

Consider the function

\[
 f(t) = f_1(t) + f_2(t) + f_3(t) = 0.3 \cos(3t) + \cos(t) + 2 \cos(0.1t),
 t \in [-60, 60],
\]

where \( f_i \) are defined in the obvious way and order. Running the EMD on \( f \) yields three IMFs \( y_i, i=1,2,3 \); the three components are (almost) perfectly recovered, in the sense that \( y_i \approx f_i, i=1,2,3 \). However, applying EMD to \( g = y_1 + y_2 \) returns \( g \) as the single IMF of the expansion, failing to separate \( y_1 \) and \( y_2 \); All these functions are shown in Fig. 1.

This experiment demonstrates that, at least this widely used and popular implementation of the EMD is not OO-consistent either, since \( y_3 \) acts as a “catalyst” between \( y_1 \) and \( y_2 \). The phenomenon in questions, however, will turn out to be an inherent feature of the EMD, independent of the envelope used, as we are about to see.

### 4.3. Causes of EMD nonlinearity

The reason OO-consistency was found not to hold in the example above is that \( g(t) = y_1(t) + y_2(t) = 0.3 \cos(3t) + \cos(t) \) contains several local “almost” extrema, indicated by the pronounced “flattening” of \( g \) around value 0 (see bottom row of Fig. 1): the reader can verify that \( g \) would, in fact, contain local extrema if the coefficient of \( f_1 \) were higher than 1/3 (as seen in the proof of Theorem 1). The superposition of the slowly varying \( y_{3}/f_1 \) is enough to make these almost extrema stand out as actual extrema (see top right in Fig. 1, where a local minimum appears near \(- 17.5 \) and a local maximum near \(- 17 \)): as some of these extrema have the wrong sign (in our example, the local maximum is negative), the envelope “picks them up” and eventually recovers \( y_{1}/f_1 \) successfully, as a result. Given that such almost extrema are, in fact, inflection points, the significance of retaining the envelope on them as well, as proposed in [24], becomes clear.

We can be reasonably certain that this mechanism of revealing hidden “almost” extrema applies on a wide range of functions, and that it does not rely on the specific envelope used. On the other hand, such “almost” extrema
are borderline cases which are not very likely to appear very frequently, so OO-consistency will hold “most of the time” (the reader is invited to tweak the coefficients in our example to observe that this is indeed the case).

Thus, the reliance of the envelope on the location and values of local extrema, which depend on the function in a fundamentally nonlinear manner, appears to be the major cause of EMD nonlinearity. In this sense, our quantitative analysis confirms the excellent qualitative analysis presented in [33], which reaches the same conclusion. It should be emphasized that is a fundamental fact holding for any reasonable envelope, rather than a side-effect of a particular envelope, and it is due to the definition of the envelope through the local extrema, highly nonlinear quantities themselves. In other words, it will not be possible to linearize EMD through a suitable envelope choice (assuming this was our ultimate hope/intention).

Additionally, two more causes of nonlinearity are mentioned in [33], namely the exact interpolation algorithm used for the construction of the envelope, along with the stopping criteria used for the iterations, which by their very nature are envelope dependent, so that their effect can potentially be minimized and even eliminated. A manifestation of these two causes are the interval endpoint effects (visible in Fig. 1, bottom left in particular; see also the discussion in Section 2.4).

5. Conclusion and summary

The main obstacle in the way of a solid theoretical foundation of EMD is its implementation-oriented definition, which makes it difficult to decide not only which parts of the implementation need to be parts of a formal definition as well, but also which of the observed features of the EMD are inherent to the algorithm, as opposed to being implementation-dependent. After a brief overview of the EMD algorithm, in which we focused, in particular, on its empirical aspects that need to be rigorously proved,
we formally described it in terms of an operator and investigated briefly its domain and range. Subsequently, constructing a counterexample, we proved that the EMD is nonlinear. Recasting the property of linearity in empirical terms, we formulated the equivalent definition of input–output consistency. Relaxing this, we arrived at the property of output–output consistency, which again through a counterexample, we showed that the EMD does not satisfy either. A detailed analysis of this counterexample identified the dependence of the envelope on the local extrema of the function (highly nonlinear quantities themselves) as the key cause of EMD nonlinearity.

Acknowledgments

The authors would like to thank the reviewers of a version of this work previously submitted for publication for their extensive and detailed comments. This material is based on works supported by the Science Foundation Ireland under Grant No. 05/Y12/1677,06/M1/006 (Claude Shannon Institute) and 08/RFP/MT1164.

References


