AN OPTIMIZATION BASED EMPIRICAL MODE DECOMPOSITION SCHEME

By

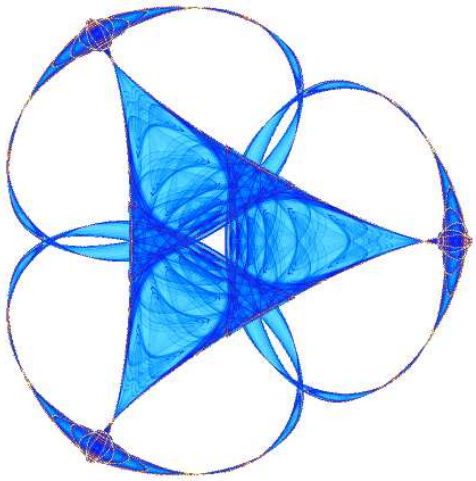
Boqiang Huang

and

Angela Kunoth

IMA Preprint Series # 2389

(February 2012)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612-624-6066 Fax: 612-626-7370

URL: http://www.ima.umn.edu

An Optimization Based Empirical Mode Decomposition Scheme

Boqiang Huang* Angela Kunoth†

Abstract

The empirical mode decomposition (EMD) has been developed by N.E. Huang *et al.* in 1998 as an iterative method to decompose a nonlinear and nonstationary univariate function additively into multiscale components. These components called intrinsic mode functions (IMFs) are constructed such that they are approximately orthogonal to each other with respect to the L_2 inner product. Moreover, the components allow for a definition of instantaneous frequencies through complexifying each component by means of the application of the Hilbert transform. This approach via analytic signals, however, does not guarantee that the resulting frequencies of the components are always non-negative and, thus, physically meaningful, and that the amplitudes can be interpreted as envelopes.

In this paper, we formulate an optimization problem which takes into account important features desired of the resulting EMD. Specifically, we propose a data-adapted iterative method which minimizes in each iteration step a smoothness functional subject to inequality constraints involving the extrema. In this way, our method constructs a sparse data-adapted basis for the input function as well as a mathematically stringent envelope for the function. Moreover, we present an optimization-based normalization to extract instantaneous frequencies from the analytic function approach. We present corresponding algorithms together with several examples.

Key words: Empirical mode decomposition (EMD), sparse data-adapted basis, intrinsic mode functions (IMFs), instantaneous frequencies, envelope, convex optimization.
AMS subject classification (MSC2000): 65Dxx, 65K10.

1 Introduction

The empirical mode decomposition (EMD) scheme introduced in [14] is a data-adapted iterative method. It decomposes a real-valued signal or univariate function f, given as discrete input data or time series of cardinality N on a finite interval $[0,T] \subset \mathbb{R}$, pointwise into additive components with multiscale features, i.e.,

$$f(t) = \sum_{j=1}^{J} g_j(t) + r_{J+1}(t) \quad \text{for every } t \in [0,T]. \tag{1.1}$$

These components, denoted as intrinsic mode functions (IMFs), have *in the ideal case* the following

*Corresponding author: bhuang@math.uni-paderborn.de, Institut für Mathematik, Universität Paderborn, Warburger Str. 100, 33098 Paderborn, Germany
†kunoth@math.uni-paderborn.de, www2.math.uni-paderborn.de/ags/kunoth

Properties 1.1 *(i) They are orthogonal to each other with respect to the $L_2(0,T)$ inner product, i.e., $(g_j, g_\ell)_{L_2(0,T)} := \int_0^T g_j(t)g_\ell(t)dt = 0$ for $j \neq \ell$;*

(ii) the local maxima are positive, the local minima are negative, and the amount of maxima and zeros differ at most by one, and the same for the minima;

(iii) the cubic spline $u_j(t)$ interpolating all the local maxima is an upper envelope *for $g_j(t)$, i.e., $g_j(t) \leq u_j(t)$ on the whole interval $[0,T]$; accordingly, the cubic spline $v_j(t)$ interpolating all the local minima is a* lower envelope *for $g_j(t)$, i.e., $g_j(t) \geq v_j(t)$ and all $t \in [0,T]$;*

(iv) the median*, defined pointwise as the mean value of the upper and lower envelope, $m_j(t) := \frac{1}{2}(u_j(t) + v_j(t))$, $t \in [0,T]$, vanishes on all of $[0,T]$;*

(v) the residual (or trend) r_{J+1} is monotone, or has at most one maximum or minimum;

(vi) the number J describing the amount of components is relatively small compared to the number of input data N;

(vii) each of the g_j exhibits an instantaneous frequency, i.e., a frequency varying over time (the precise definition will be given below).

Remark 1.2 *As we will see below in detail, the EMD scheme is an iterative method whose resulting EMD depends on different parameters. Thus, the orthogonality (i) is only expected to hold approximately. We discuss the issue of the pointwise inequalities in the envelope property (iii) in subsection 2.2.*

Recall that a cubic spline is a globally continuously twice differentiable function consisting of piecewise polynomials of degree three. Although the method works, in principle, for any kind of signal, the target objects in [14] have been measurement signals which exhibit multiscale features and which may be classified as nonlinear and nonstationary. For us, a *nonlinear* and *nonstationary* data set is a set of real-valued time series data $\{(t_\ell, f_\ell)\}_{\ell=1}^{\infty}$ if there exists a number $m \in \mathbb{Z}$ such that the common probability distribution of $f_\ell, \ldots, f_{\ell+m}$ depends on the time index ℓ. Typically, measurement data from physical processes satisfy this condition, e.g., radar, speech and many electrophysiological signals. It will be convenient in the sequel to consider a piecewise continuous function $f : \mathbb{R} \to \mathbb{R}$ instead of processing discrete time series data. Note that a (piecewise) continuous function can always be generated from discrete data by spline interpolation or by a least-squares approximation.

In view of Properties 1.1, the EMD is considered by many to be superior to Fourier or wavelet decompositions which both work with an a-priori-defined basis with prescribed frequencies. In addition, as we will see below, the iterative method generating the EMD principally does not require data given at regularly spaced times like for the standard versions of the Fast Fourier or the Fast Wavelet Transforms. By now, the EMD has steadily gained popularity in the data analysis community, and numerous modifications have been proposed to remedy some of its deficiencies, see, e.g., [4, 9, 10, 16, 17, 22, 31].

Despite its immediate application in various disciplines [13, 15, 27], one disadvantage of the original method is that its definition of instantaneous frequencies, see Properties 1.1(vii), through the Hilbert transform does not guarantee that these are non-negative and, thus, 'physically meaningful'; for a discussion of this notion, see [16]. Since this is an important point for the development of our algorithms, we will first discuss two definitions of 'instantaneous frequencies' thoroughly.

In the original paper [14], instantaneous frequencies are defined as follows: Starting from the decomposition (1.1) into real-valued intrinsic mode functions $g_j(t)$, one computes their so-called analytical representation by applying the Hilbert integral transform $\mathcal{H}$ to each $g_j(t)$. In signal

processing, this is a long-established technique going back to [5, 23], see, e.g. [3, 7]. Recall from e.g. [30] that the *Hilbert transform* $\mathcal{H}[f]$ of a function $f : \mathbb{R} \to \mathbb{R}$ is defined as the integral transform

$$\mathcal{H}[f](t) := \frac{1}{\pi}\,\mathrm{PV}\int_{-\infty}^{\infty}\frac{f(u)}{t-u}\,du := \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{-\infty}^{t-\varepsilon}\frac{f(u)}{t-u}\,du + \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{t+\varepsilon}^{\infty}\frac{f(u)}{t-u}\,du. \tag{1.2}$$

The analytic representation of the $g_j(t)$ results in complex-valued objects

$$z_j(t) := g_j(t) + i\,\mathcal{H}[g_j](t) =: a_j(t)\exp(i\theta_j(t)) \quad \text{for each } j = 1, \ldots, J, \tag{1.3}$$

with time-dependent amplitudes $a_j(t)$ and phases $\theta_j(t)$. From these, we can define corresponding instantaneous frequencies as the derivative of the phase, i.e.,

$$\omega_j(t) := \theta_j'(t), \tag{1.4}$$

see also [27]. In view of (1.1), this means that we can rewrite the signal as

$$\begin{aligned} f(t) &= \sum_{j=1}^{J}\mathrm{Re}\,(g_j(t) + i\,\mathcal{H}[g_j](t)) + r_{J+1}(t) \\ &= \mathrm{Re}\left(\sum_{j=1}^{J} a_j(t)\exp\left(i\int_{\mathbb{R}}\omega_j(t)dt\right)\right) + r_{J+1}(t), \quad t \in [0,T]. \end{aligned} \tag{1.5}$$

As desired, the representation (1.5) furnishes for each component index j an amplitude $a_j(t)$ as well as a frequency $\omega_j(t)$ depending on time. Comparing this with the Fourier representation of the signal,

$$f(t) = \mathrm{Re}\left(\sum_{j=1}^{J} a_j\,\exp(i\omega_j t)\right) + \tilde{r}_{J+1}(t), \quad t \in [0,T], \tag{1.6}$$

one observes that the Fourier components only have constant amplitude a_j and frequency ω_j. In this sense, the EMD provides a generalized Fourier representation which is particularly appropriate for nonlinear and nonstationary data.

Based on the decomposition (1.5), one can define the Hilbert Amplitude Spectrum as the set

$$H(t,\omega) := \{a_j(t) \text{ on the curves } \{t, \omega_j(t)\},\ t \in [0,T],\ j = 1, \ldots, J\}, \tag{1.7}$$

see the bottom left graphic in Figure 8 below for an illustration.

Although it provides an elegant definition of an instantaneous frequency, the analytical representation method using the Hilbert transform has some drawbacks. The most serious one is that it implicitly assumes that

$$\mathcal{H}[a(t)\cos(\theta(t))] = a(t)\,\mathcal{H}[\cos(\theta(t))]. \tag{1.8}$$

As pointed out already in [1], this is not valid in general unless the Fourier spectra of the amplitude $a(t)$ and carrier $\cos(\theta(t))$ overlap. A more fundamental difficulty is that even if $a(t) = 1$, $\mathcal{H}[\cos(\theta(t))] = \sin(\theta(t))$ is not true for an arbitrary function $\theta(t)$ [21].

According to [18], p. 115, the notion of an instantaneous, time-varying frequency may be subject to discussion. The following example from [18] may motivate our definition below, as an alternative to the classical one in (1.4). The so-called cosine modulation $f(t) = a\,\cos(\omega_0 t + \theta_0) =:$

$a\cos(\theta(t))$ for constant real a, ω_0 has the phase $\theta(t)$. Accordingly, we can derive its frequency as the derivative of the phase, $\theta'(t) = \omega_0$, which is a constant here. The idea to generalize this to allow for time-dependent frequencies is now the following: for given $f(t)$, find a representation

$$f(t) = \sum_{j=1}^{J} a_j(t)\,\cos(\theta_j(t)) + r_{J+1}(t), \quad t \in [0,T], \tag{1.9}$$

with a residual r_{J+1} satisfying Properties 1.1(v), time-dependent amplitudes $a_j(t) \geq 0$ for all $t \in [0,T]$ and *carrier functions* $\cos(\theta_j(t))$ involving monotonically increasing phases $\theta_j(t)$. If the latter is satisfied for each j, we can define their respective *instantaneous frequencies* as

$$\omega_j(t) := \theta_j'(t), \qquad t \in [0,T], \tag{1.10}$$

which are by definition non-negative and, thus, 'physically meaningful' [16]. The requirement of the amplitudes $a_j(t)$ in (1.9) to be non-negative throughout $[0,T]$ is also a reasonable assumption to achieve a certain symmetry of the IMF components around the horizontal axis.

The problem with the representation (1.9) is that it is not unique, as the following simple example shows.

Example 1.3 *Let ω_1 and ω_2 be two constants satisfying $\omega_1 > \omega_2 > 0$. Then the function $f(t)$ defined as*

$$f(t) := (1 + \cos(2\pi\,\omega_2\,t))\,\cos(2\pi\,\omega_1\,t), \qquad t \in [0,T], \tag{1.11}$$

(depicted for $\omega_1 = 1, \omega_2 = 0.3$ in the top left graphic in Figure 6) can also be expressed as

$$f(t) = 0.5\,\cos(2\pi(\omega_1+\omega_2)\,t) + \cos(2\pi\,\omega_1\,t) + 0.5\,\cos(2\pi(\omega_1-\omega_2)\,t). \tag{1.12}$$

In the representation (1.11), we have one term ($J = 1$) whereas (1.12) exhibits three additive components. The question is, therefore, which decomposition the EMD would generate. From different point of views, this issue was discussed in [24, 32].

Another simple example is

Example 1.4

$$f(t) := \cos(2\pi\,t) + a\,\cos(2\pi\,\omega\,t), \qquad t \in [0,T], \tag{1.13}$$

where a is a real positive number and $\omega \in]0,1[$. Here we have two additive components; the question is whether it is possible to find another representation with a single term but with a frequency varying over time, e.g., of the form $f(t) = a(t)\,\cos(\theta(t))$. For $a = 0.4$ and $\omega = 0.8$, the function defined in (1.13) is shown in the top graphic on the left in Figure 5.

The remainder of this paper is structured as follows. In subsection 2.1, we briefly recall initially the original EMD scheme from [14]. In subsection 2.2, we propose a novel optimization-based EMD scheme, abbreviated as OEMD method. In particular, our method guarantees by construction through the solution of a minimization problem subject to inequality constraints that the signal itself is contained in its envelope, i.e., that (iii) of Properties 1.1 holds. We present a number of numerical studies in subsection 2.3. In Section 3, we return to the definition of instantanous frequencies through the Hilbert transform and propose a new normalization which is also optimization-based, to extract these frequencies. The performance of our algorithm is substantiated by a number of experiments. Section 4 finally contains a conclusion and an outlook.

2 EMD schemes

2.1 Original EMD

According to their authors, the empirical mode decomposition method by Huang *et al.* proposed in [14] can deal specifically with data from nonlinear and nonstationary processes. The decomposition is based on the realistic assumption that any data consists of different simple intrinsic modes of oscillations which are represented by IMFs with Properties 1.1. This new method is data-adaptive in the sense that the outcome (1.1) is a representation in an a-posteriorily-defined basis, meaning that it is based on and derived from the data; in fact, in view of Properties 1.1(vi), the data is *sparse* in this basis. This basis is obtained from the data by means of an algorithm containing two nested loops: an outer one indexed by j for the generation of the jth IMF, and an inner one called *sifting process* indexed by subindex i which is based on repeated interpolation of extrema and subtraction of the median. Although the scheme is actually applied to the discrete time series data, we describe it here more conveniently in terms of a (piecewise) continuous function $f(t)$, $t \in [0, T]$.

Algorithm EMD
Step 1. Initialize $r_1(t) = f(t)$ (the residual) and $j = 1$ (index number of IMF);
Step 2. extract the j-th IMF:

(a) initialize $h_{j_0}(t) = r_j(t)$, $i = 0$;
(b) extract local minima and maxima of $h_{j_i}(t)$;
(c) compute upper and lower envelope $u_{j_i}(t)$ and $v_{j_i}(t)$ by interpolating local maxima and local minima of $h_{j_i}(t)$, respectively;
(d) compute the median $m_{j_i}(t) = \frac{1}{2}(u_{j_i}(t) + v_{j_i}(t))$;
(e) update $h_{j_{i+1}}(t) = h_{j_i}(t) - m_{j_i}(t)$ and $i = i + 1$;
(f) calculate stopping criterion SD_{j_i} according to (2.1) below;
(g) repeat steps (b) to (f) until $\mathrm{SD}_{j_i} \le \mathrm{SD}_{\mathrm{Thr}}$, define j-th IMF $g_{j_i}(t) := h_{j_i}(t)$;

Step 3. update residual $r_{j+1}(t) = r_j(t) - g_{j_i}(t)$;
Step 4. repeat Steps 2 to 3 with $j = j + 1$ until number of extrema in $r_j(t)$ is less than 2 or expected index number of IMF is met, i.e., $j = J$.

In Step 2(c), typical applications employ cubic spline interpolation, which we also do here and which gives in the opinion of many practitioners somewhat more satisfying results than interpolation by piecewise linears. With cubic spline interpolation, however, one faces the question how to deal with missing data at the boundary and how to interpret different results when using natural or artificially generated boundary conditions, see, e.g., [13]. The parameter $\mathrm{SD}_{\mathrm{Thr}}$ appearing in Step 2(g) is a user-defined quantity. Our experience is that this is a quite sensitive parameter [27]. The stoppage criterion is determined by using a Cauchy type convergence test involving the normalized difference in the $L_2(0, T)$-norm between the resulting functions of two successive sifting steps,

$$\mathrm{SD}_{j_i} := \frac{\|h_{j_{i-1}} - h_{j_i}\|_{L_2(0,T)}}{\|h_{j_{i-1}}\|_{L_2(0,T)}}. \tag{2.1}$$

The first component $g_1(t)$ generated by Algorithm EMD usually contains the finest spatial scale in the signal. The next residual $r_2(t)$, generated by subtraction from the signal $f(t)$ the first IMF $g_1(t)$ now contains information about larger scales. Repeating this process, we find by superposition of all the IMFs an additive reconstruction of the data in the form (1.1).

2.2 Optimization based EMD (OEMD)

The generation of envelopes as defined in Properties 1.1(iii) which satisfy the inequalities *pointwise on the whole interval* $[0,T]$ is not automatically guaranteed by the interpolation procedure in Step 2(c), as we will describe in the following. We have motivated already in Section 1 that we seek to find an EMD defined as in (1.9), $f(t) = \sum_{j=1}^{J} a_j(t)\cos(\theta_j(t)) + r_{J+1}(t), t \in [0,T]$, i.e., the j-th IMF $g_j(t)$ is assumed to be of the form $a_j(t)\cos(\theta_j(t))$ with $a_j(t) \geq 0$ for $t \in [0,T]$. Since $|\cos(\theta_j(t))| \leq 1$ for any t, we have $|a_j(t)\cos(\theta_j(t))| \leq |a_j(t)|$. Thus, it is naturally to realize the upper and lower envelope of the j-th IMF as $a_j(t)$ and $-a_j(t)$, respectively, and they will coincide with the IMF at the positions of the extrema where $\cos(\theta_j(t)) = \pm 1$.

Algorithm EMD realizes in Step 2(c) the computation of the upper and lower envelope by a cubic spline interpolation of the local maxima or local minima, respectively. However, the resulting 'envelope' sometimes intersects the signal; this over-/undershooting is a very undesired feature. In view of Remark 1.2, this is an artefact of the method for the construction of the envelope. Fig. 1 illustrates this phenomenon.

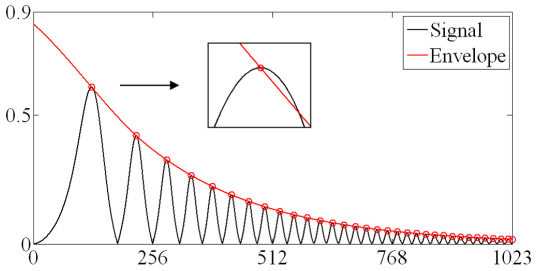


Figure 1: Intersection of the signal with the upper 'envelope' constructed by cubic spline interpolation of the local maxima.

To avoid these undesired features, we propose a construction of envelopes which strictly enforces the upper and lower inequalities in Properties 1.1(iii) at all t. Inspired by [9, 19, 28], we formulate a convex optimization problem which replaces Step 2(c) in Algorithm EMD by the solution of one such optimization problem for the upper and the lower envelope each. The optimization will be based on the minimization of a quadratic smoothness functional $S(\cdot)$ subject to pointwise inequality constraints. The functional may be chosen as the square of a Sobolev norm, or seminorm of n-th order, i.e., as the square of the $L_2(0,T)$ norm of the n-th (weak) derivative of the function v to be minimized like, e.g., in [10, 29],

$$S(v) := \|v^{(n)}\|^2_{L_2(0,T)}. \tag{2.2}$$

For discrete signals, the computation of such functionals is, in fact, realized by taking differences of function values so that the actual smoothness of the function to be minimized is not relevant.

Thus, for the generation of the jth IMF, in the ith sifting step, denote by t_k^+ and t_k^- the positions of the local maxima and minima, respectively, of the intermediate function $h_{j_i}(t)$. We then solve the optimization problems

$$\begin{array}{lll} \text{(P1)} & \text{Minimize} & S(X) \quad \text{over all } X(t),\ t \in [0,T], \\ & \text{subject to} & h_{j_i}(t) \leq X(t) \text{ for all } t \in [0,T], \\ & \text{and} & h_{j_i}(t_k^+) = X(t_k^+), \quad \text{all } k; \end{array}$$

$$\begin{array}{lll} \text{(P2)} & \text{Minimize} & S(X) \quad \text{over all } X(t),\ t \in [0,T], \\ & \text{subject to} & h_{j_i}(t) \geq X(t) \text{ for all } t \in [0,T], \\ & \text{and} & h_{j_i}(t_k^-) = X(t_k^-), \quad \text{all } k. \end{array}$$

The minimizers of these problems which are by construction upper and lower envelopes for each $t \in [0, T]$ are denoted by $u_{j_i}(t)$ and $v_{j_i}(t)$, respectively. We replace now Step 2(c) in Algorithm EMD by the following step and call the resulting scheme Algorithm OEMD (Optimization-based EMD). The quadratic smoothness functional $S(\cdot)$ is supposed to be fixed beforehand for the whole algorithm.

Algorithm OEMD

Step 2(c) Solve optimization problem (P1) with maxima $h_{j_i}(t_k^+)$, denote minimizer by $u_{j_i}(t)$;

Solve optimization problem (P2) with minima $h_{j_i}(t_k^-)$, denote minimizer by $v_{j_i}(t)$.

The optimization problems (P1) and (P2) are both quadratic programming problems: the smoothness functional $S(\cdot)$ was chosen to be convex, and the inequality and equality constraints are linear. Such problems can be solved by the algorithms suggested in [2]. We have used here the CVX tool box for convex programming problems from [6]. The problem is solved by the interior point method SeDuMi: the inequalities are appended to the quadratic functional by means of a barrier function, and the conditions for optimality lead to a coupled system of nonlinear equations where in each iteration step Newton's method is applied.

2.3 Numerical studies

We want to illustrate the performance of the new OEMD method in comparison with Algorithm EMD for several synthetic examples. To distinguish between the two different results, the ones from the EMD are marked by a superscript E while the ones from the OEMD are marked by a superscript O. The implementation has been performed using MATLAB 7.12.0.635 (R2011a).

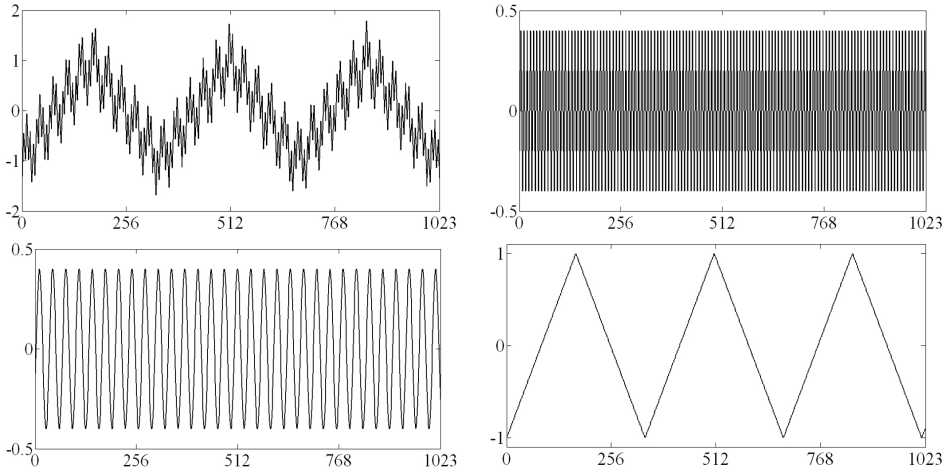


Figure 2: Signal f (top left), additively composed from sine wave g_2, bottom left, and two piecewise linear continuous functions g_1, top right, and, g_3, bottom right, with different periodicities.

The first example is a synthetic signal f on the time interval $[0, 1023]$ proposed in [25]. It consists of three components: a sinusoid g_2 with period $p_2 = 100/3$ superimposed by two triangular-shaped waveforms g_1, g_3 with a smaller and larger period than p_2 ($p_1 = 8, p_3 = 338$), respectively. The signal together with its three components are illustrated in Fig. 2.

First we display in Fig. 3 the results of OEMD with smoothness functional involving the third derivative, i.e., $n = 3$ in (2.2), in comparison with the ones from [25]. We observe that OEMD

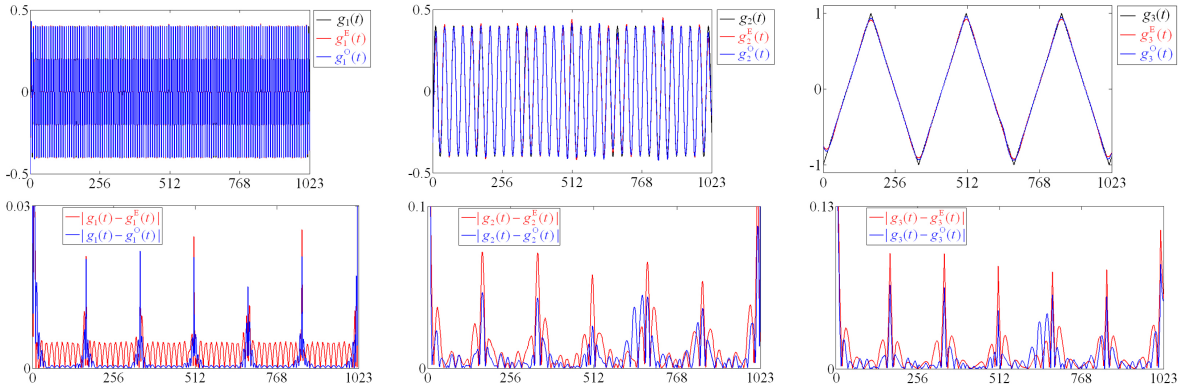


Figure 3: Test signal from Fig. 2. Comparison of the results from OEMD with $n = 3$ in the smoothness functional (2.2) with EMD from [25]. Top row: display of the three components, g_j (black), g_j^{E} (red), g_j^{O} (blue), after the final respective sifting step, for $j = 1, 2, 3$ from left to right. Bottom row: absolute differences of $|g_j(t) - g_j^{\mathrm{E}}(t)|$, $t \in [0, T]$ (red), and $|g_j(t) - g_j^{\mathrm{O}}(t)|$, $t \in [0, T]$ (blue).

qualitatively performs better, particularly concerning the absolute errors. Next we study for this example the effect of changing the parameter n in the smoothness functional (2.2). Fig. 4 shows the absolute differences between the original components $g_j(t)$ and the reconstructed $g_j^{\mathrm{E}}(t)$ or $g_j^{\mathrm{O}}(t)$ for $n = 1, 2, 3$. As expected, the best results are obtained for $n = 3$, i.e, requiring minimal variation of the second derivative of the reconstructed IMFs. Therefore, we will for all subsequent experiments fix (2.2) for $n = 3$.

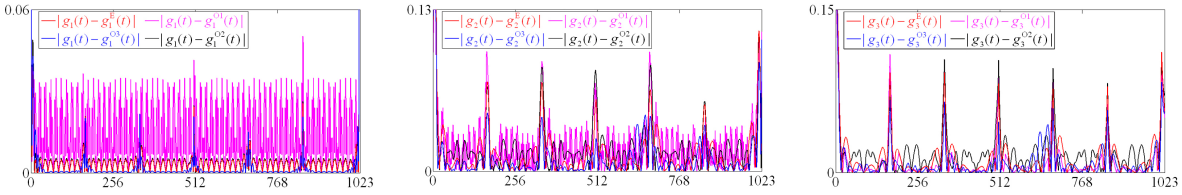


Figure 4: Absolute differences between original and generated IMFs using EMD from [25] ($|g_j(t) - g_j^{\mathrm{E}}(t)|$, $t \in [0, T]$, red) and OEMD for $n = 1$ ($|g_j(t) - g_j^{\mathrm{O1}}(t)|$, $t \in [0, T]$, magenta), $n = 2$ ($|g_j(t) - g_j^{\mathrm{O2}}(t)|$, $t \in [0, T]$, black) and $n = 3$ ($|g_j(t) - g_j^{\mathrm{O3}}(t)|$, $t \in [0, T]$, blue).

Next we want to give a first answer to the non-uniqueness problem raised in Example 1.4. Fig. 5 shows the decomposition results given by OEMD, again compared to the EMD scheme in [24]. Here, $a = 0.4$ and $\omega = 0.8$. For these particular parameters, it was claimed in [24] that the EMD cannot separate the two components (1.13) into two frequencies with a fixed sifting number $i = 10$. In the top row, the middle graphic shows that EMD generates slightly symmetric envelopes in the first sifting process. Both envelopes cross the signal almost 30 times. The right one shows that OEMD forms the envelope under a strict mathematical definition and creates an oscillatory median which will contribute in the sifting process for separating the second component from the synthetic signal.

A similar decomposition result is illustrated in Fig. 6 towards the non-uniqueness problem in Example 1.3 for $\omega_1 = 1$ and $\omega_2 = 0.3$. In the top row, the EMD also generates nearly symmetric envelopes, and both envelopes intersect the signal nearly 41 times. In the bottom row, OEMD almost separates all three frequencies from the synthetic signal in the representation with three additive components (1.12) while the EMD fails in the separation. Together with this paper, we

provide an online animation to present the detailed behavior of both EMD and OEMD during the sifting process on these two non-uniqueness problems.

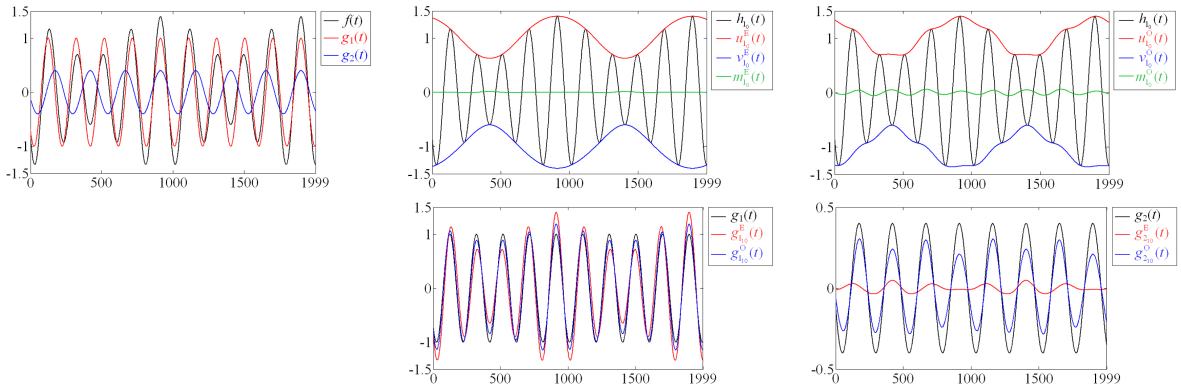


Figure 5: Function (1.13) for $a = 0.4$ and $\omega = 0.8$. Top row, left graphic: signal f (black) and its two components g_1 (red) and g_2 (blue); middle and right graphics: intermediate function with its upper and lower envelopes and median in first sifting process given by EMD: h_{1_0} (black), $u^{\mathrm{E}}_{1_0}$ (red), $v^{\mathrm{E}}_{1_0}$ (blue) and $m^{\mathrm{E}}_{1_0}$ (green) and generated by OEMD h_{1_0} (black), $u^{\mathrm{O}}_{1_0}$ (red), $v^{\mathrm{O}}_{1_0}$ (blue) and $m^{\mathrm{O}}_{1_0}$ (green). Bottom row: two original and computed IMFs g_j (black), $g^{\mathrm{E}}_{j1_0}$ (red) and $g^{\mathrm{O}}_{j1_0}$ (blue).

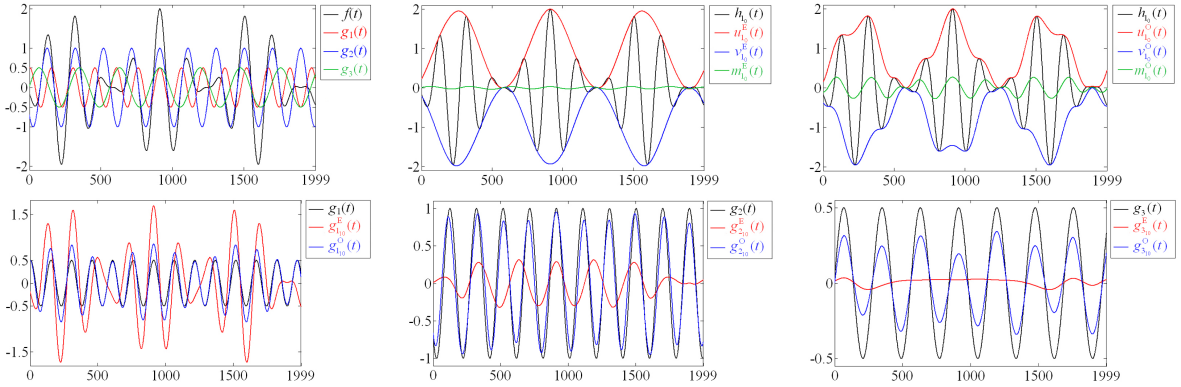


Figure 6: Signal f from (1.12) for $\omega_1 = 1$ and $\omega_2 = 0.3$. Top row, left graphic: f (black) and its three components g_1 (red), g_2 (blue) and g_3 (green); middle and right graphics: intermediate functions with their upper and lower envelopes and medians in first sifting process given by EMD: h_{1_0} (black), $u^{\mathrm{E}}_{1_0}$ (red), $v^{\mathrm{E}}_{1_0}$ (blue), and $m^{\mathrm{E}}_{1_0}$ (green) and generated by OEMD h_{1_0} (black), $u^{\mathrm{O}}_{1_0}$ (red), $v^{\mathrm{O}}_{1_0}$ (blue) and $m^{\mathrm{O}}_{1_0}$ (green). Bottom row: three original and computed IMFs g_j (black) $g^{\mathrm{E}}_{j1_0}$ (red) and $g^{\mathrm{O}}_{j1_0}$ (blue).

Remark 2.1 *We want to make a few remarks on the arithmetic complexity of our algorithm. Both OEMD and the original EMD have in common that one needs to repeatedly solve an interpolation problem. For a representation in cubic B-splines, this means to solve a linear system with a sparse system matrix in N variables where N is the amount of samples representing the functions, i.e., the system matrix has $\mathcal{O}(N)$ entries. In the j_ith Step 2(c) of Algorithm EMD or OEMD, K_i interpolation conditions are posed. The system matrix contains as its entries the B-splines evaluated at t_k^+ for (P1) or at t_k^- for (P2). Since B-splines are by construction compactly supported and we are dealing with one-dimensional problems, the only nonzero entries of this system matrix are clustered around and parallel to the diagonal. Depending on the grid spacing relative to the location to the maxima or minima, a typical form would be a pentadiagonal*

matrix. Again the properties of B-splines ensure that this matrix can be decomposed in a stable way by LU factorization without pivoting which requires optimal $\mathcal{O}(N)$ arithmetic operations.

In all the examples considered here, the functions are sampled at $N = 1024$ times on a uniform grid with grid spacing 2^{-10}.

Solving an optimization problem as in OEMD is more expensive than simply such an interpolation; in fact, one has to expect that the interior point method requires operations of the order $\mathcal{O}(N \log N)$. We believe that, as an alternative, this could be accelerated to again $\mathcal{O}(N)$ operations by using a primal-dual active set method, see, e.g., [8, 20]. This method is an inner-outer iterative process where in the outer step the points where the inequality conditions are active are identified, followed by an inner step to only solve the optimality system consisting of coupled linear equations for which also no longer Newton's method would be required.

Remark 2.2 *Our optimization approach does not yet guarantee monotonically increasing phases $\theta_j(t)$ as envisaged in (1.9). Also a certain relation between the smoothness of the amplitudes for different j's and in relation to the phases may be necessary to assume similar as in [4]. This point will be the subject of further studies.*

3 Optimization based normalization for instantaneous Fourier analyses

In this section, we return to the analytical signal method presented in Section 1 and address the question of the practical computation of instantaneous frequencies. In fact, recalling the representation (1.3), they could be computed as in (1.4) [1, 21].

An alternative method was proposed in [11,16] to seperate any IMF empirically and uniquely into an envelope (AM) and a unity-valued carrier (FM) part. Here 'unity-valued carrier' means that the upper/lower envelope of the carrier should be equal to ± 1 for any $t \in [0, T]$. In other words, the 'unity-valued carrier' is an approximation to the ideal carrier $\cos(\theta_j(t))$ in (1.9).

An example of the normalization procedure described below is provided in Fig. 7. The scheme

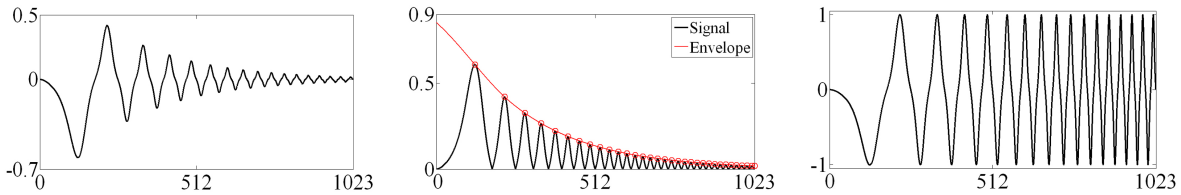


Figure 7: Example of normalization scheme. Left: given IMF g_j; middle: $|g_j(t)|$ for all $t \in [0, T]$ and 'approximate' envelope, i.e, cubic spline interpolating maxima of modulus of g_j; right: normalized signal $\tilde{g}_j$ according to (3.3).

is as follows. Given an IMF g_j, identify all local maxima of the modulus $|g_j(t)|$, $t \in [0, T]$ of g_j. Next, interpolate these local maxima by a cubic spline denoted as u_j^{E}, see the middle graphic of Fig. 7. The spline interpolation of the local maxima of the modulus of g_j will guarantee that the normalized g_j will be symmetric with respect to the horizontal axis. In view of the remarks at the beginning of subsection 2.2, this envelope is not one in a mathematically strict sense; we call it 'empirical' envelope. By construction, it is uniquely defined. This will be used now to normalize the IMF g_j according to

$$\tilde{g}_j(t) := \frac{g_j(t)}{u_j^{\mathrm{E}}(t)}, \quad t \in [0, T], \tag{3.3}$$

Ideally, all extrema of $\tilde{g}_j$ should now have value ± 1. By Taylor expansion of both carrier and its *quadrature*, defined as a 90° shift of the phase angle of the carrier, one can approximate the instantaneous frequency at any discrete time t_k by

$$\theta_j'(t_k) \approx \left| \frac{\cos(\theta_j(t_{k+1})) - \cos(\theta_j(t_{k-1}))}{(t_{k+1} - t_{k-1})\sqrt{1 - \cos^2(\theta_j(t_k))}} \right| \approx \left| \frac{\tilde{g}_j(t_{k+1}) - \tilde{g}_j(t_{k-1})}{(t_{k+1} - t_{k-1})\sqrt{1 - \tilde{g}_j^2(t_k)}} \right|. \tag{3.4}$$

see [11,16]. Note that, according to (3.4), the right hand side is not well-defined at those points in which $\cos(\theta_j(t_k)) = 1$ ($\tilde{g}_j(t_k) = 1$). Thus, one cannot use this formula around those points. Instead, one may use an interpolation of the instantaneous frequency based on those points in which $\cos(\theta_j(t_k)) \neq 1$ ($\tilde{g}_j(t_k) \neq 1$), e.g. $\cos(\theta_j(t_k)) < 0.9$ ($\tilde{g}_j(t_k) < 0.9$) [11].

Unfortunately, since cubic interpolation may cause overshooting or undershooting, see section 2.2, this normalization may not always give a unity-valued carrier which may result in a negative $1 - \tilde{g}_j^2(t_k)$ in (3.4). Reference [16] suggests to excute the normalization iteratively until a 'good' unity-valued carrier is found. To overcome this problem, we propose an optimization-based normalization method as follows.

Algorithm Normalization
Step 1: For a given IMF $g_j(t)$, define $f_j(t) = |g_j(t)|$, $t \in [0, T]$ and identify maxima of $f_j(t)$;
Step 2: solve optimization problem (P1) and get envelope function $u_j^{\mathrm{O}}(t)$ of $f_j(t)$;
Step 3: normalize IMF by $\tilde{g}_j(t) = g_j(t)/u_j^{\mathrm{O}}(t)$, $t \in [0, T]$;
Step 4: obtain instantaneous frequency by (3.4).

Since this optimization-based method enables us to obtain an envelope in a strict mathematical sense (see Properties 1.1(iii)), our normalization is completed after one sweep. To illustrate this procedure, an example is shown in Fig. 8. Here, a signal with variable sampling frequency is chosen to compare the performance between the proposed optimization based-method and the Hilbert transform based method from [11,16]. The signal as a continuous function is defined as

$$f(t) = \sin(8\pi\,(\sin(2\pi\,t) + t)), t \in [0, 1], \tag{3.5}$$

which is sampled on a grid with non-uniform grid spacing as follows. The sampling frequency is set as 1024 Hz in the interval [0, 0.5] and as 512 Hz in the interval [0.5, 1]. In Fig. 8, the first row shows the original signal and its decomposition by employing a B-spline based EMD method from [26]. The second row shows the amplitude spectrum based on the Hilbert space method and our proposed method. Note that there is a clearly visible oscillation in the Hilbert amplitude spectrum defined according to (1.7) at the position where the sampling frequency changes, see the graphic at the bottom left of Fig. 8. This can be explained by the deficiency of a correct estimation of instantaneous frequencies in the Hilbert transform setting. There is no such phenomenon in the amplitude spectrum based on our Algorithm Normalization which is displayed in the bottom right graphic of Fig. 8.

Finally, the normalization scheme has three important consequences: first, and most importantly, the unity-valued carrier enables us to directly compute quadrature, see the sentence after (3.3) for this definition. Second, the carrier has unity amplitude; therefore, it satisfies the Bedrosian identity (1.8) automatically. Third, the normalization method changes the measure of error considered in Nuttall's theorem in [21] to the error between the approximated quadrature and the real quadrature when one wants to describe a time-frequency spectrum towards a signal. Combined with the second consequence, it can be concluded that the unity-valued carrier may enable us to provide a sharper local energy than that from Hilbert transform.

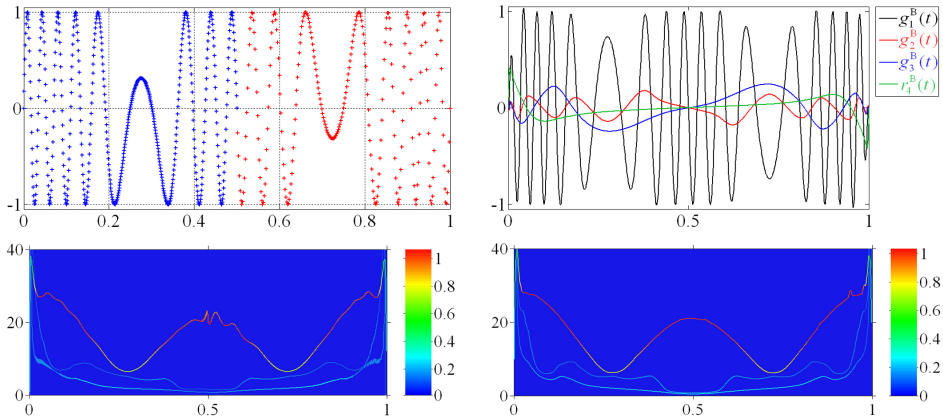


Figure 8: Comparison between the traditional Hilbert transform based method for the instantaneous frequency estimation and Algorithm Normalization. Top row: original signal sampled on different grids (left) and its decomposition results from [26] (right); Bottom row: amplitude spectrum based on the Hilbert transform based method (1.7) (left) and our proposed method (right).

4 Conclusion and Outlook

We have introduced an optimization based EMD (OEMD) method for one-dimensional signals. The main feature of our algorithm is that we construct an envelope for the signal in a strict mathematical sense by formulating and solving convex optimization problems. By construction, our method does not create over- or undershootings which are disadvantageous side effects of traditional EMD spline interpolation methods. The numerical experiments show that we obtain a somewhat more separate decomposition than the original EMD method [14, 24]. We also proposed an optimization-based normalization scheme to extract instantaneous frequencies in the traditional Hilbert transform based method and demonstrated superior features for a synthetically generated signal.

Some extensions of our scheme to the two-dimensional case with an application to images is presented in [12]. Once the decomposition is achieved which is likely to improve the one in, e.g., [17], one could apply the monogenic Clifford algebra-valued techniques described there to construct a generalized Hilbert-Huang transform for images.

Acknowledgments

Boqiang Huang receives funding as a PostDoc of the Alexander von Humboldt Foundation. Angela Kunoth's research was supported in part by the Institute for Mathematics and its Applications (IMA) at the University of Minnesota with funds provided by the National Science Foundation (NSF).

References

[1] E. Bedrosian, A product theorem for Hilbert transforms, Proc. IEEE 51 (1963) 868-869.

[2] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge, U.K., available online: http://www.stanford.edu/∼boyd/cvxbook/, 2004.

[3] L. Cohen, Time-Frequency Analysis, Prentice Hall PTR, U.S.A, 1994.

[4] I. Daubechies, J. Lu, H.-T. Wu, Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool, Appl. Comput. Harmon. Anal. 30 (2011) 243-261.

[5] D. Gabor, Theory of communication. Part 1: The analysis of information, in: Part III: Radio and Communication Engineering, J. Inst. Electr. Eng. 93 (1946) 429-441.

[6] M. Grant, S. Boyd, CVX Users Guide, available online: http://cvxr.com, 2011.

[7] S.L. Hahn, Hilbert Transforms in Signal Processing, Artech House, U.S.A, 1996.

[8] H. Hoffmann, A. Kunoth, P. Strack, Fast multiscale PDAS solvers for quadratic optimization problems involving elliptic PDEs under inequality constraints, in preparation.

[9] T.Y. Hou, Z. Shi, Adaptive data analysis via sparse time-frequency representation, Adv. Adapt. Data Anal. 3 (2011) 1-28.

[10] T.Y. Hou, Z. Shi, Sparse time-frequency representation of multiscale data by nonlinear matching pursuit, Manuscript, 2011, submitted for publication.

[11] T.Y. Hou, M.P. Yan, Z. Wu, A variant of the EMD method for multi-scale data, Adv. Adapt. Data Anal. 1 (2009) 483-516.

[12] B. Huang, An optimization based empirical mode decomposition scheme for images, in preparation.

[13] N.E. Huang, S.S.P. Shen, Hilbert-Huang Transform and Its Applications, World Scientific Publishing, Singapore, 2005.

[14] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zhang, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. A 454 (1998) 903–995.

[15] N.E. Huang, M.-L. Wu, W. Qu, S.R. Long, S.S.P. Shen, Applications of Hilbert-Huang transform to non-stationary financial time series analysis, Appl. Stochastic Models Bus. Ind. 19 (2003) 245-268.

[16] N.E. Huang, Z. Wu, S.R. Long, K.C. Arnold, X. Chen, K. Blank, On instantaneous frequency, Adv. Adapt. Data Anal. 1 (2009) 177-229.

[17] G. Jager, R. Koch, A. Kunoth, R. Pabel, Fast empirical mode decompositions of multivariate data based on adaptive spline-wavelets and a generalization of the Hilbert-Huang-Transform (HHT) to arbitrary space dimensions, Adv. Adapt. Data Anal. 2 (2010) 337-358.

[18] S. Mallat, A Wavelet Tour of Signal Processing, third ed.: The Sparse Way, Academic Press, U.S.A, 2008.

[19] J. Mattingley, S. Boyd, Real-time convex optimization in signal processing, IEEE Signal Process. Mag. 27 (2010) 50-61.

[20] J. Nocedal, S.J. Wright, Numerical Optimization, 2nd ed., Springer, 2006.

[21] A.H. Nuttall, On the quadrature approximation to the Hilbert transform of modulated signals, Proc. IEEE 54 (1966) 1458-1459.

[22] S. Peng, W.-L. Hwang, Null space pursuit: An operator-based approach to adaptive signal separation, IEEE Trans. Signal Process. 58 (2010) 2475-2483.

[23] v.d. Pol, The fundamental principles of frequency modulation, in: Part III: Radio and Communication Engineering, J. Inst. Electr. Eng. 93 (1946) 153-158.

[24] G. Rilling, P. Flandrin, One or two frequencies? The empirical mode decomposition answers, IEEE Trans. Signal Process. 56 (2008) 85-95.

[25] G. Rilling, P. Flandrin and P. Gonçalves, On empirical mode decomposition and its algorithms, Proc. IEEE-EURASIP Workshop Nonlinear Signal Image Process. NSIP-03, Grado, Italy, 2003.

[26] J. Rudi, Empirical Mode Decomposition via Adaptiver Wavelet-Approximation (in German), Diploma Thesis, Institut für Mathematik, Universität Paderborn, Germany, 2010.

[27] J. Rudi, R. Pabel, G. Jager, R. Koch, A. Kunoth, H. Bogena, Multiscale analysis of hydrologic time series data using the Hilbert-Huang-Transform (HHT), Vadose Zone J. 9 (2010) 925-942.

[28] G. Sell, M. Slaney, Solving demodulation as an optimization problem, IEEE Trans. Audio Speech Lang. Process. 18 (2010) 2051-2066.

[29] W. Stefan, R. Renaut, A. Gelb, Improved total variation-type regularization using higher-order edge detectors, SIAM J. Imaging Sci. 3 (2010) 232-251.

[30] E.C. Titchmarsh, Introduction to the Theory of Fourier Integral, second ed., Oxford University Press, 1948.

[31] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adapt. Data Anal. 1 (2009) 1-41.

[32] H.-T. Wu, P. Flandrin, I. Daubechies, One or two frequencies? The synchrosqueezing answers. Adv. Adapt. Data Anal. 3 (2011) 29-39.