Chapter 4: Time-Frequency Analysis, Adaptive Representation of Nonlinear and Non-stationary Signals

Wavelet Analysis and Applications
Tao Qian, Mang I. Vai and Xu Yuesheng, Eds.
Applied and Numerical Harmonic Analysis, 287–297

Time-Frequency Aspects of Nonlinear Fourier Atoms

Qiuhui Chen, Luoqing Li and Tao Qian

Abstract. In the standard Fourier analysis one uses the linear Fourier atoms $\{e^{int} : n \in \mathbb{Z}\}$. With only the linear phases nt Fourier analysis can not expose the essence of time-varying frequencies of nonlinear and non-stationary signals. In this note we study time-frequency properties of a new family of atoms $\{e^{in\theta_a(t)} : n \in \mathbb{Z}\}$, non-linear Fourier atoms, where a is any but fixed complex number with $|a| < 1$, and $d\theta_a(t)$ a harmonic measure on the unit circle parameterized by t. The nonlinear Fourier atoms $\{e^{in\theta_a(t)} : n \in \mathbb{Z}\}$ were first noted in [12] with some examples and theoretically studied in [8]. In this note we show that the real parts $\cos\theta_a(t), |a| < 1$, form a family of intrinsic mode functions introduced in the HHT theory [5]. We prove that for a fixed a the set $\{e^{in\theta_a(t)} : n \in \mathbb{Z}\}$, constitutes a Riesz basis in the space $L^2([0, 2\pi])$. Some miscellaneous results including Shannon type sampling theorems are obtained.

Mathematics Subject Classification (2000). Primary 94A12; Secondary 42C15.

Keywords. Nonlinear Fourier atom; Hilbert-Huang transform; time-frequency analysis; sampling theorem.

1. Introduction

The frequency of non-stationary signals varies with time. The traditional Fourier analysis, however, can not expose the time-varying property of frequency of non-stationary signals. This is due to the basic fact that in Fourier analysis a general signal is superposition of harmonic waves of which each has a constant frequency.

Recently Norden E. Huang [5] presented a new time-frequency method for nonlinear and non-stationary signal analysis: Hilbert-Huang Transform (HHT).

Qiuhui Chen is supported in part by NSFC under grant 10201034 and the Project-sponsored by SRF for ROCS, SEM. Luoqing Li is supported in part by NSFC under grant 10371033. Tao Qian is supported by University of Macau under research grant RG065/03-04S/QT/FST and Macao Science and Technology Development Fund (FDCT) 051/2005/A.

By using the algorithm of Empirical Mode Decomposition(EMD), any multi-component can be decomposed into a finite sum of *intrinsic mode functions* (IMFs), which are essentially of mono-component. Based on the EMD decomposition, further treatments can be done under the framework of the ideal time-frequency analysis. The notion of IMF defined by Huang plays a crucial role in the HHT theory. The original definition of IMFs is an engineering description: The occurrences of local maximums and minimums take turn, and between a pair of adjacent local extremes, the signal is monotone and passes through the zero once, and is of the local symmetry, i.e. the mean of a upper and an adjacent lower envelope based on the local extremes, respectively, is of the zero value. Experiments show that IMFs behave nicely with Hilbert transformation and offer meaningful instantaneous frequencies. IMFs are considered to be essentially of mono-component discussed in the notion of instantaneous frequency [2]. Based on the Bedrosian [1] and Nuttall [6] theorems a natural question occurs: For an amplitude - frequency modulation (AM and FM) signal $f = a(t)\cos\theta(t)$, under what conditions on a and θ the associated quadrature signal $a(t)e^{i\theta(t)}$ becomes analytic?

In [8], Qian proves that a strictly increasing function $\theta(t), t \in [0, 2\pi]$ with $m(\theta([0, 2\pi])) = 2\pi$ gives rise to an analytic signal $e^{i\theta(t)}$ if and only if $d\theta(t)$ is a harmonic measure on the circle, and this result has a counterpart for strictly increasing functions $\Theta(s)$ with $m(\Theta(\mathbb{R})) = 2\pi$ on the whole real line. In this note, we explore some time-frequency aspects of the family of the new nonlinear Fourier atoms $\{e^{in\theta_a(t)} : n \in \mathbb{Z}\}, |a| < 1$, where $d\theta_a(t)$ is a harmonic measure, that is, the derivative of $\theta_a(t)$ is the Poisson kernel (see §2). For further development of the theme the reader is recommended to [10] and [11].

In section 2, we show that $\cos\theta_a(t)$ is of mono-component. That essentially means that $\theta_a'(t) > 0$, the Hilbert transform of $\cos\theta_a(t)$ is $\sin\theta_a(t)$, and $\theta_a(t)$ can be decomposed into a sum of a linear part and a nonlinear but periodic part. In section 3, we prove that $\{e^{in\theta_a(t)} : n \in \mathbb{Z}\}$ forms a Riesz basis of $L^2([0, 2\pi])$, with the Riesz bounds $\frac{1-|a|}{1+|a|}$ and $\frac{1+|a|}{1-|a|}$. Section 4 mainly concerns sampling theorems in relation to this new family of atoms.

Before we start there is one point to note. Functions studied in below sometimes are defined on the unit circle, or equivalently on intervals with length 2π with equal function values at the two ends of each of such intervals; and sometimes are defined on the whole real line but 2π-periodic. These two cases should be clearly distinguished, especially when dealing with certain operators acting on function spaces. For instance, images of Fourier transformation operator on the first type of functions are Fourier series, and the involved integral regions in that case are intervals of length 2π. Correspondingly, when we deal with the first type of functions, then the Hilbert transformation is $\widetilde{\mathcal{H}}$ (see §2); and for the second type, the Hilbert transformation is $\mathcal{H}$. In order to avoid the ambiguity we shall clearly note which case we are working with, but we do not assign special notations for the two types of functions.

2. Nonlinear Fourier Atoms and Behavior with Hilbert Transform

We start our discussion with the Möbius transformation

$$\tau_a(z) = \frac{z-a}{1-\overline{a}z}$$

that is a conformal mapping one to one and onto from the unit disc to itself with the condition $\tau(a) = 0$. Define $\theta_a(t)$ by

$$e^{i\theta_a(t)} = \tau_a\left(e^{it}\right) = \frac{e^{it}-a}{1-\overline{a}e^{it}}.$$

Note that θ_a is defined on the unit circle and its derivative is the Poisson kernel (see [4] or [8])

$$\theta_a'(t) = p_a(t) = \frac{1-|a|^2}{1-2|a|\cos(t-t_a)+|a|^2}.$$

The function θ_a may be continuously extended to the whole real line with the property $\theta_a(t+2\pi) = \theta_a(t) + 2\pi$ whose derivative $p_a(t)$ is continuous and 2π-periodic. In this section, unless otherwise stated, we shall treat θ_a as defined in the whole real line. The corresponding period functions $e^{i\xi\theta_a(t)}$, $\xi > 0$, except for the trivial case $a = 0$ corresponding to $e^{i\xi t}$ of the linear phase ξt, are not included in the general form of Picinbono [7]. Indeed, the derivatives of the phases of the signals in [7] are sums of the Poisson kernels on the real line and therefore not periodic. The atomic case of Picinbono was studied in [8].

Let $a = |a|e^{it_a}$. Then

$$e^{i\theta_a(t)} = \frac{e^{it}-a}{1-\overline{a}e^{it}} = \frac{e^{it}-|a|e^{it_a}}{1-|a|e^{-it_a}e^{it}} = \frac{1-|a|e^{i(t_a-t)}}{1-|a|e^{i(t-t_a)}}\,e^{it} = \frac{A(t)}{\overline{A(t)}}\,e^{it},$$

where $A(t) = 1-|a|e^{i(t_a-t)}$. By noting that

$$\mathrm{Arg}A(t) = \arctan\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)},$$

we get the explicit expression

$$\theta_a(t) = t + 2\arctan\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)}. \tag{2.1}$$

Note that the first part is linear and the second part is periodic, and such a decomposition is unique.

It is interesting that the signal $\cos\theta_a(t)$ is a mono-component with frequency modulation. To see this, we need to show that its Hilbert transform is the corresponding sine function $\sin\theta_a(t)$.

The circular Hilbert transform ([4]) of a function $f = \sum_{k\in\mathbb{Z}} c_k e^{ikt} \in L^2([0,2\pi])$ is defined by

$$\widetilde{\mathcal{H}}f(t) = -i\sum_{k\in\mathbb{Z}}\mathrm{sgn}(k)c_k e^{ikt}.$$

It has a singular integral expression

$$\widetilde{\mathcal{H}}f(t) = \text{v.p.}\frac{1}{2\pi}\int_{-\pi}^{\pi} \cot\left(\frac{t-s}{2}\right) f(s)ds, \quad \text{a. e.}$$

On the other hand, the Hilbert transform for a function f on the real line is formally defined by

$$\mathcal{H}f(t) = \text{v.p.}\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(s)}{t-s}ds.$$

The Bedrosian Theorem says that if f is a real-valued signal of low frequencies and g is real-valued of high frequencies, then $\mathcal{H}(fg) = f\mathcal{H}g$. In accordance with the Bedrosian Theorem, it has been well accepted that if $\mathcal{H}\left(a(t)\cos\phi(t)\right) = a(t)\sin\phi(t)$ and $\phi'(t) \geq 0$, then meaningful instantaneous amplitudes and frequencies may be defined through the amplitude-frequency modulation signal $s(t) = a(t)\cos\phi(t)$. In the case we regard $s(t)$ as of mono-component [2]. The following theorem states that $\cos\theta(t)$ is a mono-component with constant amplitude.

Theorem 2.1. *(i) Treating* $\cos\theta_a(t)$ *as a function defined on the unit circle, we have*

$$\widetilde{\mathcal{H}}\cos\theta_a(t) = \sin\theta_a(t);$$

and (ii) Treating $\cos\theta_a(t)$ *as a* 2π*-periodic function on the whole real line, we have*

$$\mathcal{H}\cos\theta_a(t) = \sin\theta_a(t).$$

Proof. The proof of (i) is contained in [8] and [12]. To make the current paper self-contain we present an outline of the proof. Note that $\tau_a(z)$ is an analytic function and $e^{i\theta_a(t)}$ its boundary value. We therefore have

$$e^{i\theta_a(t)} = \sum_{k=0}^{+\infty} c_k e^{ikt}$$

for a fast decay sequence $\{c_k : k \in \mathbb{Z}\}$. By the definition of the circular Hilbert transform, we have

$$\widetilde{\mathcal{H}}\left(e^{i\theta_a(t)}\right) = -i\left(\sum_{k=1}^{+\infty} c_k e^{ikt}\right) = -i\left(e^{i\theta_a(t)} - c_0\right).$$

This concludes $\widetilde{\mathcal{H}}\cos\theta_a(t) = \sin\theta_a(t)$.

Now we prove (ii). First note that in the Riemann improper integral and principal value sense the Hilbert transformation is well defined for the oscillatory function $\cos\theta_a(t)$. By the definition of the Hilbert transform, we have

$$\begin{aligned}\text{p.v.}\ \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1}{x-t}\cos\theta_a(t)dt &= \text{p.v.}\ \frac{1}{\pi}\int_0^{2\pi}\sum_{k=-\infty}^{\infty}\frac{1}{x-t+2k\pi}\cos\theta_a(t)dt \\ &= \text{p.v.}\ \frac{1}{2\pi}\int_0^{2\pi}\cot\left(\frac{x-t}{2}\right)\cos\theta_a(t)dt,\end{aligned}$$

where we used the identity $\lim_{N\to\infty}\sum_{k=-N}^{N}\frac{1}{x-t+2k\pi}=\frac{1}{2}\cot\left(\frac{x-t}{2}\right)$ which may be found, for instance, from [9]. Then the conclusion of (i) implies that of (ii). The proof of Theorem 2.1 is complete. □

We offer an explicit representation for the function $\cos\theta_a(t)$. First, by (2.1), we get that

$$\begin{aligned}\cos\theta_a(t) &= \cos\left(t+2\arctan\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)}\right)\\ &= \cos\left(2\arctan\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)}\right)\cos t\\ &\quad -\sin\left(2\arctan\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)}\right)\sin t.\end{aligned}$$

Second, using the formulas $\cos 2t=\frac{1-\tan^2 t}{1+\tan^2 t}$ and $\sin 2t=\frac{2\tan t}{1+\tan^2 t}$, we have

$$\begin{aligned}\cos\theta_a(t) &= \frac{1-\left(\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)}\right)^2}{1+\left(\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)}\right)^2}\cos t-\frac{2\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)}}{1+\left(\frac{|a|\sin(t-t_a)}{1-|a|\cos(t-t_a)}\right)^2}\sin t\\ &= \frac{\left(1-|a|\cos(t-t_a)\right)^2-\left(|a|\sin(t-t_a)\right)^2}{\left(1-|a|\cos(t-t_a)\right)^2+\left(|a|\sin(t-t_a)\right)^2}\cos t\\ &\quad -\frac{2|a|\sin(t-t_a)\left(1-|a|\cos(t-t_a)\right)}{\left(1-|a|\cos(t-t_a)\right)^2+\left(|a|\sin(t-t_a)\right)^2}\sin t.\end{aligned}$$

Finally, through a direct computation, we have

$$\begin{aligned}\cos\theta_a(t) &= \frac{\cos t\left(1-|a|\cos(t-t_a)\right)^2-\cos t\left(|a|\sin(t-t_a)\right)^2}{\left(|a|\sin(t-t_a)\right)^2+\left(1-|a|\cos(t-t_a)\right)^2}\\ &\quad -\frac{2|a|\sin t\sin(t-t_a)\left(1-|a|\cos(t-t_a)\right)}{\left(|a|\sin(t-t_a)\right)^2+\left(1-|a|\cos(t-t_a)\right)^2}\\ &= \frac{\cos t\left(1-2|a|\cos(t-t_a)+|a|^2\cos 2(t-t_a)\right)}{1+|a|^2-2|a|\cos(t-t_a)}\\ &\quad -\frac{2|a|\sin t\sin(t-t_a)+|a|^2\sin t\sin 2(t-t_a)}{1+|a|^2-2|a|\cos(t-t_a)}\\ &= \frac{\cos t-2|a|\left(\cos t\cos(t-t_a)+\sin t\sin(t-t_a)\right)}{1+|a|^2-2|a|\cos(t-t_a)}\end{aligned}$$

$$+\frac{|a|^2\left(\cos t\cos 2(t-t_a)+\sin t\sin 2(t-t_a)\right)}{1+|a|^2-2|a|\cos(t-t_a)}$$
$$=\frac{\cos t-2|a|\cos t_a+|a|^2\cos(t-2t_a)}{1+|a|^2-2|a|\cos(t-t_a)}.$$

In particular, when a is a real number less than 1, $\cos\theta_a(t)$ can be simplified into

$$\cos\theta_a(t)=\frac{\left(1+|a|^2\right)\cos t-2|a|}{1+|a|^2-2|a|\cos t}.$$

3. Nonlinear Fourier Atoms as Riesz Basis

With the notation $\theta_a'=p_a$, $p_a(t)$ being the Poisson kernel, we have $p_0=1$. There hold the estimates for p_a :

$$\frac{1-|a|}{1+|a|}\le p_a(t)\le\frac{1+|a|}{1-|a|}.$$

Define

$$L^2_{p_a}([0,2\pi])=\left\{f:[0,2\pi]\to\mathbb{C}\ :\ \int_0^{2\pi}|f(t)|^2p_a(t)dt<\infty\right\}.$$

It is a Hilbert space equipped with the inner product

$$\langle f,g\rangle_{p_a}=\int_0^{2\pi}f(t)\overline{g(t)}p_a(t)dt$$

and the norm

$$\|f\|_{p_a}=\left(\int_0^{2\pi}|f(t)|^2p_a(t)dt\right)^{\frac{1}{2}}.$$

Note that for $a=0$ the space $L^2_{p_a}([0,2\pi])$ reduces to the classic case $L^2([0,2\pi])$ with the norm $\|f\|=\left(\int_0^{2\pi}|f(t)|^2dt\right)^{\frac{1}{2}}$. We have the equivalence between the two norms

$$\sqrt{\frac{1-|a|}{1+|a|}}\,\|\cdot\|\le\|\cdot\|_{p_a}\le\sqrt{\frac{1+|a|}{1-|a|}}\,\|\cdot\|. \tag{3.1}$$

Note that all the function sets $L^2_{p_a}([0,2\pi])$, $|a|<1$, are identical with different but equivalent norms. Through change of variable the classical Carleson's Theorem reduces to the assertion (also see [8]) that for any $f\in L^2_{p_a}([0,2\pi])$,

$$f(t)=\sum_{n\in\mathbb{Z}}c_n^a(f)e^{in\theta_a(t)},\quad a.e.$$

The identity of the function sets then implies that the last equality also holds for functions in $f\in L^2([0,2\pi])$. That is, the standard square integrable functions can be approximated by the nonlinear Fourier atoms with the weighted Fourier coefficients $c_n^a(f)$.

Theorem 3.1. *For any fixed a the system* $\{e^{in\theta_a(t)} : n \in \mathbb{Z}\}$ *forms a Riesz basis of* $L^2([0, 2\pi])$ *with the upper bound* $\frac{1+|a|}{1-|a|}$ *and the lower bound* $\frac{1-|a|}{1+|a|}$, *i.e.*

$$\frac{1-|a|}{1+|a|}\| \{c_j\} \|^2 \ \leq \ \left\| \sum_{j\in\mathbb{Z}} c_j e^{ij\theta_a(t)} \right\|^2 \ \leq \ \frac{1+|a|}{1-|a|}\| \{c_j\} \|^2 \tag{3.2}$$

for any ℓ^2*- sequence* $\{c_j : j \in \mathbb{Z}\}$.

Proof. Note that $\{\frac{1}{\sqrt{2\pi}} e^{in\theta_a(t)} : n \in \mathbb{Z}\}$ is a weighted orthonormal system with the weight p_a. This implies the Plancherel identity:

$$\|f\|_{p_a}^2 = \sum_{j\in\mathbb{Z}} \left|c_j^a(f)\right|^2,$$

where $f \in L^2_{p_a}([0, 2\pi])$ and the Fourier coefficients $c_j^a(f) = \left\langle f, \frac{1}{\sqrt{2\pi}} e^{ij\theta_a(t)} \right\rangle_{p_a}$. The Plancherel identity then implies that for any sequence $\{c_j : j \in \mathbb{Z}\} \in \ell^2$ the function

$$f(t) := \sum_{j\in\mathbb{Z}} c_j e^{ij\theta_a(t)}, \quad t \in [0, 2\pi],$$

is well defined and belongs to $L^2_{p_a}([0, 2\pi])$, and, further for $j \in \mathbb{Z}$, $c_j^a(f) = c_j$. Since f also belongs to $L^2([0, 2\pi])$, the equivalence (3.1) of the two norms of f together with the Plancherel identity in $L^2_{p_a}([0, 2\pi])$ then gives the relation (3.2). □

4. Shannon Type Sampling Theorem

For any function $f \in L^2(\mathbb{R})$, the generalized Fourier transform is defined by

$$\widehat{f}^a(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\xi\theta_a(t)} p_a(t) dt, \tag{4.1}$$

where both θ_a and p_a are 2π-periodic extensions of the corresponding functions on the unit circle. Note that when $a = 0$, it is the classic Fourier transform

$$\widehat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\xi t} dt.$$

In the distribution sense, we can check that the generalized Fourier transform of $\cos\theta_a(t)$ is $\frac{1}{2}(\delta(\xi - 1) + \delta(\xi + 1))$. The inverse formula of (4.1) is

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \widehat{f}^a(\xi) e^{i\xi\theta_a(t)} d\xi. \tag{4.2}$$

Theorem 4.1. *Suppose that* $\mathrm{supp}\widehat{f^a} \subset [-\Omega, \Omega]$. *Then*

$$f(t) = \sum_{n\in\mathbb{Z}} f\left(\theta_a^{-1}(n\frac{\pi}{\Omega})\right) \frac{\sin(\Omega\theta_a(t) - n\pi)}{(\Omega\theta_a(t) - n\pi)}. \tag{4.3}$$

Proof. Since $\widehat{f^a}$ has compact support in $[-\Omega, \Omega]$, we can represent $\widehat{f^a}$ by its standard Fourier series

$$\widehat{f^a}(\xi) = \sum_{n\in\mathbb{Z}} c_n e^{-in\xi\frac{\pi}{\Omega}},$$

where

$$c_n = \frac{1}{2\Omega}\int_{-\Omega}^{\Omega} \widehat{f^a}(\xi) e^{in\xi\frac{\pi}{\Omega}} d\xi = \frac{1}{2\Omega}\int_{-\infty}^{\infty} \widehat{f^a}(\xi) e^{in\xi\frac{\pi}{\Omega}} d\xi.$$

By invoking the inverse formula (4.2), we have

$$\begin{aligned} c_n &= \frac{\sqrt{2\pi}}{2\Omega}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \widehat{f^a}(\xi) e^{i\xi\theta_a\left(\theta_a^{-1}(n\frac{\pi}{\Omega})\right)} d\xi \\ &= \frac{\sqrt{2\pi}}{2\Omega} f\left(\theta^{-1}(n\frac{\pi}{\Omega})\right). \end{aligned}$$

It then follows that

$$\begin{aligned} f(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \widehat{f^a}(\xi) e^{i\xi\theta_a(t)} d\xi \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\Omega}^{\Omega} \sum_{n\in\mathbb{Z}} \frac{\sqrt{2\pi}}{2\Omega} f\left(\theta^{-1}(n\frac{\pi}{\Omega})\right) e^{i\xi\left(\theta_a(t) - n\frac{\pi}{\Omega}\right)} d\xi. \end{aligned}$$

Since the series for $\widehat{f^a}$ converges also in $L^1([0, 2\pi])$, interchanging the order of integration and summation leads to

$$\begin{aligned} f(t) &= \sum_{n\in\mathbb{Z}} f\left(\theta^{-1}(n\frac{\pi}{\Omega})\right)\int_{-\Omega}^{\Omega} \frac{1}{2\Omega} e^{i\xi\left(\theta_a(t) - n\frac{\pi}{\Omega}\right)} d\xi \\ &= \sum_{n\in\mathbb{Z}} f\left(\theta_a^{-1}(n\frac{\pi}{\Omega})\right) \frac{\sin\left(\Omega\theta_a(t) - n\pi\right)}{\Omega\theta_a(t) - n\pi}. \end{aligned}$$

□

The convergence speed of the new sampling theorem can also be improved through over-sampling. To this end, we choose an arbitrary function g such that $\operatorname{supp}\widehat{g} \subset [-\delta, \delta]$ with $\delta = l\Omega$ for some positive integer $l > 1$, and $\widehat{g} = 1$ on $[-\Omega, \Omega]$. Thus, $\widehat{f^a}\widehat{g} = \widehat{f^a}$. Note that $\widehat{f^a}$ has the Fourier series expansion $\widehat{f^a} = \sum_{n\in\mathbb{Z}} c_n e^{in\xi\frac{\pi}{\delta}}$ with $c_n = \frac{\sqrt{2\pi}}{2\delta} f\left(\theta_a^{-1}\left(n\frac{\pi}{\delta}\right)\right)$. Hence

$$\begin{aligned} f(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \widehat{f^a}(\xi) e^{i\xi\theta_a(t)} d\xi \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\delta}^{\delta} \widehat{f^a}(\xi)\widehat{g}(\xi) e^{i\xi\theta_a(t)} d\xi \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \int_{-\delta}^{\delta} \sum_{n\in\mathbb{Z}} \frac{\sqrt{2\pi}}{2\delta} f\left(\theta_a^{-1}\left(n\frac{\pi}{\delta}\right)\right) e^{-in\xi\frac{\pi}{\delta}} \widehat{g}(\xi) e^{i\xi\theta_a(t)} d\xi \\
&= \sum_{n\in\mathbb{Z}} f\left(\theta_a^{-1}\left(n\frac{\pi}{\delta}\right)\right) \frac{1}{\sqrt{2\pi}} \int_{-\delta}^{\delta} \frac{\sqrt{2\pi}}{2\delta} \widehat{g}(\xi) e^{i\xi\left(\theta_a(t) - n\frac{\pi}{\delta}\right)} d\xi \\
&= \sum_{n\in\mathbb{Z}} f\left(\theta_a^{-1}\left(n\frac{\pi}{\delta}\right)\right) \frac{\sqrt{2\pi}}{2\delta} g\left(\theta_a(t) - n\frac{\pi}{\delta}\right).
\end{aligned}
$$

If, in particular, we choose $\delta = (1+\lambda)\Omega$ and g being

$$
\widehat{g}(\xi) = \begin{cases} 1, & |\xi| \le \Omega, \\ 1 - \frac{|\xi|-\Omega}{\lambda\Omega}, & \Omega \le |\xi| \le (1+\lambda)\Omega, \\ 0, & |\xi| \ge (1+\lambda)\Omega. \end{cases}
$$

Then

$$
g(t) = \frac{4}{\sqrt{2\pi}} \frac{\sin\left(t\Omega(1+\frac{\lambda}{2})\right)\sin\left(t\frac{\Omega\lambda}{2}\right)}{\lambda\Omega t^2},
$$

and we have

$$
f(t) = \sum_{n\in\mathbb{Z}} f\left(\theta_a^{-1}\left(\frac{n\pi}{(1+\lambda)\Omega}\right)\right) G_\lambda\left(\theta_a(t) - \frac{n\pi}{(1+\lambda)\Omega}\right), \tag{4.4}
$$

where

$$
G_\lambda(x) = \frac{2}{(1+\lambda)\Omega} \frac{\sin\left((\theta_a(t) - \frac{n\pi}{(1+\lambda)\Omega})\Omega(1+\frac{\lambda}{2})\right)\sin\left((\theta_a(t) - \frac{n\pi}{(1+\lambda)\Omega})\frac{\Omega\lambda}{2}\right)}{\lambda\Omega(\theta_a(t) - \frac{n\pi}{(1+\lambda)\Omega})^2}.
$$

Note that (4.3) may be obtained by letting $\lambda \to 0$ in (4.4).

To end this note we point out that the standard Poisson summation formula may be generalized to the nonlinear Fourier atoms.

Denote by $L^2_{p_a}(\mathbb{R})$ the Hilbert space of square integrable functions on $\mathbb{R}$ with respect to the periodic weight function $p_a(t)$. Let $f \in L^2_{p_a}(\mathbb{R})$ and set

$$
\widetilde{f}(t) = \sum_{k\in\mathbb{Z}} f(t + 2k\pi).
$$

Then $\widetilde{f} \in L^2_{p_a}([0, 2\pi])$, and

$$
\widetilde{f}(t) = \sum_{k\in\mathbb{Z}} c_k^a e^{ik\theta_a(t)}.
$$

We now show that

$$
\widehat{f^a}(k) = c_k^a, \qquad k \text{ is any integer.}
$$

In fact, as in the standard case, for $k \in \mathbb{Z}$

$$\begin{aligned}
\widehat{f^a}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ik\theta_a(t)} f(t) p_a(t) dt \\
&= \frac{1}{\sqrt{2\pi}} \sum_{n\in\mathbb{Z}} \int_0^{2\pi} e^{ik\theta_a(t)} f(t+2n\pi) p_a(t) dt \\
&= \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} e^{ik\theta_a(t)} \sum_{n\in\mathbb{Z}} f(t+2n\pi) p_a(t) dt \\
&= \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} e^{ik\theta_a(t)} \widetilde{f}(t) p_a(t) dt \\
&= c_k^a.
\end{aligned}$$

We thus have $\widetilde{f}(t) = \sum_{k\in\mathbb{Z}} \widehat{f^a}(k) e^{ik\theta_a(t)}$. Let $t = t_0$, where $\theta_a(t_0) = 0$, this last relation becomes

$$\sum_{k\in\mathbb{Z}} f(t_0 + 2k\pi) = \sum_{k\in\mathbb{Z}} \widehat{f^a}(k),$$

the new Poisson formula.

References

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

[2] B. Boashash *Estimating and interpreting the instantaneous frequency of a signal, I. Fundamentals*. Proc. IEEE **80** (1992), 417-430.

[3] I. Daubechies, *Ten Lectures on Wavelets*. CBMS **61** SIAM, Philadelphia, 1992.

[4] J. B. Garnett, *Bounded Analytic Functions*. Academic Press, 1987.

[5] N. E. Huang et al, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*. Proc. R. Soc. London, **454** A (1998), 903-995.

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

[7] B. Picinbono, *On instantaneous amplitude and phase of signals*. IEEE Transaction on Signal Processing, **45** (1997), 552-560.

[8] T. Qian, *Unit analytic signals and harmonic measures*. J. Math. Anal Appl. **314** (2006), 526-536.

[9] T. Qian, *Singular integrals with holomorphic kernels and Fourier multipliers on star-shap Lipschitz curves*. Studia Mathematica, **123**(3) (1997), 195-216.

[10] T. Qian, *Characterization of Boundary Values of Functions in Hardy Spaces With Application in Signal Analysis*, Journal of Integral Equations and Applications, Volume **17** Issue 2 (Summer 2005), 159-198.

[11] T. Qian, *Mono-components for decomposition of signals*, Math. Meth. Appl. Sci. 2006; **29**:1187-1198.

[12] T. Qian, Q. H. Chen and L. Q. Li, *Unit analytic signals with nonlinear phase.* Physica D: Nonlinear Phenomena, **303** (1-2) (2005), 80-87.

Qiuhui Chen
Faculty of Mathematics and Computer Science
Hubei University
Wuhan, 430062, P. R. China
e-mail: sqh@hubu.edu.cn

Luoqing Li
Faculty of Mathematics and Computer Science
Hubei University
Wuhan, 430062, P. R. China
e-mail: lilq@hubu.edu.cn

Tao Qian
Department of Mathematics
Faculty of Science and Technology
University of Macau
Macao (via Hong Kong)
e-mail: fsttq@umac.mo

Wavelet Analysis and Applications
Tao Qian, Mang I. Vai and Xu Yuesheng, Eds.
Applied and Numerical Harmonic Analysis, 299–319

Mono-components for Signal Decomposition

Tao Qian

Abstract. In relation to the study of instantaneous frequency, HHT and the EMD algorithm in signal analysis people have been trying to find solutions of the eigenfunction problem: Find $f(t) = \rho(t)e^{i\theta(t)}$ such that $Hf = -if$, $\rho(t) \geq 0$ and $\theta'(t) \geq 0, a.e.$, where Hf is Hilbert transform of f. This article serves as a survey on some recent studies, and presents some new results as well. In the survey part we first review the systematic study on the unimodular case, and then give a detailed account on a fundamental class of non-unimodular solutions, called H-atoms, in terms of starlike functions in one complex variable. As new result we construct certain circular mono-components that do not fall into the category of H-atoms but of the form $\rho(t)e^{i\theta_a(t)}$, where $\rho(t) \geq 0$, and $e^{i\theta_a(t)}$ is some Fourier atom, as well as those of the form $\rho(s)e^{i\phi_a(s)}$, where $e^{i\phi_a(s)}$ is one on the line induced from some Fourier atom under Cayley transform.

Mathematics Subject Classification (2000). Primary: 30D55, 31A20, 31C05; Secondary: 42A50, 42B20.

Keywords. Analytic signal, instantaneous frequency, Hilbert transform, Möbius transform, mono-component, empirical mode decomposition, HHT (Hilbert-Huang transform), starlike functions.

1. Introduction

In signal analysis one has been trying to understand, for a given signal, what are its instantaneous amplitude, instantaneous phase, and instantaneous frequency. A signal, denoted by $f(t)$, is assumed to be a real-valued locally (Lebesgue) integrable function. A common approach to find the instantaneous objects is as follows. One first introduces the associated analytic signal, $Af(t) = f(t) + iHf(t)$, where Hf is

The work was supported by research grant of the University of Macau No. RG079/04-05S/QT/FST and Macao Science and Technology Development Fund 051/2005/A.

the Hilbert transform of f, being assumed to exist. Hilbert transform is formally defined by the principal value singular integral

$$Hf(t) = \text{p.v.}\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{f(s)}{t-s}ds,$$

which has the Fourier multiplier form

$$Hf(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\xi t}\left(-i\text{sgn}(\xi)\right)\hat{f}(\xi)d\xi,$$

where Fourier transform is defined by

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-i\xi t}f(t)dt,$$

and sgn is the signum function that takes value 1 if $\xi > 0$; and -1 if $\xi < 0$.

The function Af may be written in the form $Af(t) = \rho(t)e^{i\theta(t)}$, with $\rho(t) \geq 0$, a.e. Consequently,

$$f(t) = \rho(t)\cos\theta(t). \tag{1.1}$$

Due to the relation $H^2 = -I$, where I stands for the identity operator, Af satisfies the relation

$$H(Af) = -iAf, \tag{1.2}$$

which is equivalent to

$$H(\rho(\cdot)\cos\theta(\cdot))(t) = \rho(t)\sin\theta(t). \tag{1.3}$$

With the uniquely determined modulation (1.1), one calls $\rho(t)$ and $\theta(t)$ the *instantaneous amplitude* and *instantaneous phase*, respectively, provided $\theta'(t) \geq 0$, a.e. Should the requirement be met, the function $\theta'(t)$ is defined to be the *instantaneous frequency*. Unfortunately, the requirement $\theta' \geq 0$ can hardly be met, and the definitions of instantaneous amplitude, phase and frequency via the associated analytic signal Af can be erroneous.

In a related aspect, in [9], Huang proposed an algorithm, called Empirical Mode Decomposition, to decompose a signal into a sum

$$f(t) = \sum \rho_i(t)\cos\theta_i(t), \tag{1.4}$$

where each entry of the sum is expected to be a real-mono-component (see definition in §2). He also obtained numerically rapid convergence. As a new type of signal decomposition Huang's algorithm appears very effective in practice, and has attracted a wide attention. On the other hand, the algorithm suffers for it does not always result in the desired decomposition in terms of mono-components. A mathematical formulation providing precise concepts and a related approximation theory is desired.

In [13], [17], [14], [15], a systematic study on the unimodular case $\rho \equiv 1$ is carried out. In a recent paper [16] the study of the unimodular case is extend to the general non-unimodular case. We found that the well established theory of

starlike functions in one complex variable best fits our purpose. Boundary values of starlike functions provide easily accessible circular mono-components.

The writing plan of the paper is as follows. §2 contains notation and terminology related to the eigenfunction problem. It also contains a discussion on dual mono-components. In §3 we deal with unimodular solutions of the problem. We introduce Fourier atoms and finite Blaschke products. In §4 we exam the relations between the solutions and the boundary values of functions in Hardy H^p spaces. Based on the unimodular solutions in §3, through elementary conformal mappings, we construct a large class of unimodular solutions. In §5 we present an important and fundamental class of solutions, called H-atoms, that is identical with a class of starlike functions in one complex variable. This, in particular, gives rise to a large class of solutions with the non-unimodular case. In §6 we construct some new types of mono-components, that do not fall into the category of H-atoms, but of the form $\rho(t)e^{i\theta_a(t)}$, where $\rho \geq 0$ and $e^{i\theta_a(t)}$ is a Fourier atom, as well as those of the form $\rho(t)e^{i\phi_a(t)}$, where $e^{i\phi_a(t)}$ is a mono-component on the line induced from a Fourier atom under Cayley transform.

We note that, apart from §6, all the other section §2, §3 and §4 serve as a survey of the main results obtained in [17], [14], [15] and [16].

The author wishes to acknowledge his sincere thanks to Sheng Gong who kindly recommended comprehensive references in complex analysis, including starlike functions in one and several complex variables. The author wishes to thank Qiu-hui Chen and Luo-qing Li, for their collaborations in this subject, especially in [17], and in some related work in relation to wavelets ([2], [3]), and the inspiring discussions from time to time. Sincere thanks are due to Jing-xin Yin and Gui-fang Xie for their very kind and constant help in supplying the author the necessary references.

2. Notation and Terminology

Denote by $\mathbf{S}$ for $\mathbf{S} = \mathbb{D}$ or $\mathbf{S} = \mathbb{C}^+$, the earlier being the open unit disc and the latter being the upper-half complex plane. In this notation $H_{\mathbf{S}}$ stands for $H_{\mathbb{C}^+}$ or $H_{\mathbb{D}}$, where $H_{\mathbb{C}^+}$ is the standard Hilbert transformation, H, on the line, and $H_{\mathbb{D}}$ is the circular Hilbert transformation, often denoted by $\tilde{H}$, on the circle. The circular Hilbert transformation is defined through

$$\tilde{H}f(t) = \text{p.v.}\frac{1}{2\pi}\int_{-\pi}^{\pi} \cot(\frac{t-s}{2})f(s)ds$$

with the Fourier multiplier form based on the Fourier expansion of $f(t)$:

$$\tilde{H}f(t) = \sum_{k=-\infty}^{\infty}{}' -i\,\text{sgn}(k)c_k e^{ikt}, \quad f(t) = \sum_{k=-\infty}^{\infty} c_k e^{ikt}.$$

Let f be an eigenfunction of the circular or non-circular Hilbert transformation $H_{\mathbf{S}}$. Then $H_{\mathbf{S}}f = kf, k \in \mathbb{C}$. Since $H_{\mathbf{S}}^2 f = k^2 f = -f$, we obtain $k = \pm i$, where i is the complex imaginary unit.

Definition 2.1. A function f is said to be an $H_{\mathbf{S}}$-eigenfunction if $H_{\mathbf{S}}f = -if$; and a dual $H_{\mathbf{S}}$-eigenfunction if $H_{\mathbf{S}}f = if$. An $H_{\mathbf{S}}$-eigenfunction f is called an $\mathbf{S}$-mono-component if, with the form $f(t) = \rho(t)e^{i\theta(t)}$, it satisfies $\rho(t) \geq 0$ and $\theta'(t) \geq 0$, a.e.; and, a dual $H_{\mathbf{S}}$-eigenfunction f is called a dual $\mathbf{S}$-mono-component if, with the form $f(t) = \rho(t)e^{i\theta(t)}$, it satisfies $\rho(t) \geq 0$ and $\theta'(t) \leq 0$, a.e.

In the sequel, with $\mathbf{S} = \mathbb{C}^+$, with terminologies like $H_{\mathbb{C}^+}$-eigenfunctions and $\mathbb{C}^+$-mono-components, we suppress the subscript $\mathbb{C}^+$ and simply write *H-eigenfunctions* and *mono-components*, respectively; and, with $\mathbf{S} = \mathbb{D}$, we write *circular H-eigenfunctions* and *circular mono-components* for $H_{\mathbb{D}}$-eigenfunctions and $\mathbb{D}$-mono-components, etc.

Very often, we investigate $\mathrm{Re}f$ instead of f, and, with the form $f(t) = \rho(t)e^{i\theta(t)}$, we have $\mathrm{Re}f = \rho(t)\cos\theta(t)$. In the case, we have, $H_{\mathbf{S}}f = \mp if$ if and only if $H_{\mathbf{S}}(\rho(\cdot)\cos\theta(\cdot))(t) = \pm\rho(t)\sin\theta(t)$. In the case, we call $\rho(t)\cos\theta(t)$ *a real* $H_{\mathbf{S}}$*-eigenfunction*, or *a real dual* $H_{\mathbf{S}}$*-eigenfunction,* respectively. If there is no confusion, then we may omit the word "*real*", and still call it *a* $H_{\mathbf{S}}$*-eigenfunction*, or *a dual* $H_{\mathbf{S}}$*-eigenfunction*.

If a signal is not $\mathbf{S}$-mono-component or a dual $\mathbf{S}$-mono-component, then it is called a $\mathbf{S}$*-multi-component*, or simply *multi-component*. Signals are usually multi-components.

Based on this notion the task would be two-fold. The first is to establish a bank of mono- and dual mono-components. The second is to study adaptive decomposition of signals into linear combinations of mono- and dual mono-components. The present paper addresses the first. Along with the results previously obtained in [13], [17], [14], [15] and [16], in this article we construct certain solutions of the eigenfunction problem of the form $\rho(t)e^{i\theta_a(t)}$, where $\rho \geq 0$ is non-constant and $e^{i\theta_a(t)}$ is a so-called Fourier atom on the circle (see [17] and [14]), as well as certain solutions of the form $\rho(s)e^{i\phi_a(s)}$, where $e^{i\phi_a(s)}$ is induced from a Fourier atom under Cayley transform.

In below we give some remarks on dual mono-components.

When expending $f \in L^2([0, 2\pi])$ into its Fourier series

$$f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos kt + b_k \sin kt,$$

or its complex Fourier series

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{ikt},$$

the entries $\sin kt = \cos(\pi/2 - kt)$ and $e^{-ikt}, k > 0$, are dual circular mono-components. These can be verified directly, or derived from Theorem 2.2 (see below). They are also dual mono-components on the line if they are considered as periodic functions (see §3). The following result allows us to merely concentrate to the non-dual case.

Theorem 2.2. *$\rho(t)e^{i\theta(t)}$ is a (circular) mono-component if and only if $\rho(t)e^{-i\theta(t)}$ is a dual (circular) mono-component.*

Proof. Assume that $f(t) = \rho(t)e^{i\theta(t)}$ is a mono-component. We have,

$$H(\rho(\cdot)\cos\theta(\cdot))(t) = \rho(t)\sin\theta(t),$$

and, since $H^2 = -I$,

$$H(\rho(\cdot)\sin\theta(\cdot))(t) = -\rho(t)\cos\theta(t).$$

They can be re-written as

$$H(\rho(\cdot)\cos(-\theta(\cdot)))(t) = -\rho(t)\sin(-\theta(t)), \quad H(\rho(\cdot)\sin(-\theta(\cdot)))(t) = \rho(t)\cos(-\theta(t)).$$

The last two relations are equivalent to

$$H(\rho(\cdot)e^{-i\theta(\cdot)})(t) = i\rho(t)e^{-i\theta(t)}.$$

Therefore, $\rho(t)e^{-i\theta(t)}$ is a dual H-eigenfunction. Since $\rho \geq 0, -\theta' \leq 0$, it is a dual mono-component. The argument is reversible. For the circular case we replace H by $\tilde{H}$. The proof is complete. □

We show that for $k > 0$, $\sin kt$ is a dual (circular) mono-component. In fact, Theorem 2.2 implies that ie^{-ikt} is a dual (circular) mono-component. Therefore $\sin kt = \mathrm{Re}(ie^{-ikt})$ is a dual (circular) mono-component. In general, $f = u + iv$ is a dual (circular) eigenfunction if and only if $H_{\mathbf{S}}u = -v$.

In the rest of the paper we are bound to find mono-components, that is to solve the following

Hilbert Transformation Eigenfunction Problem: Find $f(t) = \rho(t)e^{i\theta(t)}$ such that $Hf = -if$, $\rho(t) \geq 0$ and $\theta'(t) \geq 0, a.e.$, where H is Hilbert transformation.

On the circle we can ask the same question with H replaced by $\tilde{H}$. We adopt the approach that we first find solutions on the unit circle and then induce the corresponding solutions on the line.

3. Unimodular Solutions of the Eigenfunction Problem

This section deals with the unimodular case $f(t) = \rho(t)e^{i\theta(t)}$, with $\rho = 1$. We proved the following theorem ([17],[14]).

Theorem 3.1. *Assume that θ is a continuous and strictly increasing function on $[0, 2\pi]$ with $|\theta([0, 2\pi])| = 2\pi$. Then the following two conditions are equivalent.*

(i) *$d\theta(t)$ is a harmonic measure on the unit circle.*

(ii)

$$\tilde{H}\cos\theta(t) = \sin\theta(t) + \mathrm{Im}a, \quad \text{and} \quad \tilde{H}\sin\theta(t) = -\cos\theta(t) - \mathrm{Re}a \tag{3.1}$$

for some $a \in \mathbb{D}$.

We note that by definition $d\theta(t)$ is a harmonic measure on the circle if $\theta'(t)$ is a Poisson kernel on the circle. For any complex number $a \in \mathbb{D}$ denote by $e^{i\theta_a(t)}$ the unimodular circular mono-component established through Theorem 3.1, called a *Fourier atom*. The function $\theta_a(t)$ is defined through the boundary value of a typical Möbius transform sending a to zero:

$$\tau_a(z) = \frac{z-a}{1-\bar{a}z}, \quad e^{i\theta_a(t)} = \frac{e^{it}-a}{1-\overline{a}e^{it}}. \tag{3.2}$$

The theorem implies that $\theta_a'(t)$ is the Poisson kernel of the circle at the point a, and $\theta_a(t)$ is, in fact, an absolutely continuous function. Note that when $a = 0$, $e^{i\theta_a(t)} = e^{it}$. The finite product of k copies of e^{it} is e^{ikt}. A generalized Fourier series and weighted Fourier transform theory are studied in [14].

This simplest unimodular case is further extended to a product of finite many Möbius transforms corresponding to Blaschke product, as given in [15].

Theorem 3.2. *Assume that θ is an absolutely continuous function on $[0, 2\pi]$ strictly increasing with $m(\theta([0, 2\pi])) = 2\pi n$, where m stands for the Lebesgue measure. Then the following two conditions are equivalent.*

(i) *$d\theta(t)$ is a sum of a number of n harmonic measures on the unit circle.*

(ii)

$$\tilde{H}\cos\theta(t) = \sin\theta(t) - (-1)^n \mathrm{Im}(\prod_{k=1}^{n} a_k) \tag{3.3}$$

and

$$\tilde{H}\sin\theta(t) = -\cos\theta(t) + (-1)^n \mathrm{Re}(\prod_{k=1}^{n} a_k) \tag{3.4}$$

for some $a_k \in D, k = 1, ..., n$.

Based on finite Blaschke products on the circle one can introduce two types of mono- and dual mono-components on the line. One is periodic extensions of the Fourier atoms. We make the extension $\theta_a(t + 2\pi) = \theta_a + 2\pi, -\infty < t < \infty$. Based on this, the periodic function $e^{i\theta_a(t)}, -\infty < t < \infty$, is a unimodular periodic mono-component on the line.

The second type is images of those functions under Cayley transformation: $\kappa : \mathbb{C}^+ \to \mathbb{D}$,

$$z = \kappa(w) = \frac{i-w}{i+w} \tag{3.5}$$

and the corresponding boundary relation

$$e^{it} = \frac{i-s}{i+s}, \quad s = \tan\frac{t}{2}. \tag{3.6}$$

Let $\phi_a(s) = \theta_a(2\arctan s), -\infty < s < \infty$, we obtain that $e^{i\phi_a(s)}$ is a unimodular non-periodic mono-component on the line ([14], [15]).

The second type was previously studied in [13] based on a different approach.

We cite the following spectrum results for the two types of mono-components ([2]). They will be recalled in §6.

Viewing $e^{i\theta_a(t)}$ as a periodic function on the line, we have ([2])

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\theta_a(t)}e^{-i\xi t}dt = -\sqrt{2\pi}a\delta(\xi) + \frac{\sqrt{2\pi}(1-|a|^2)}{\bar{a}}\sum_{k=1}^{\infty}\bar{a}^k\delta(\xi-k). \quad (3.7)$$

For the non-periodic type on the line we have ([2])

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{i\phi_a(t)}e^{-i\xi t}dt = -\sqrt{2\pi}\delta(\xi) + \frac{2\sqrt{2\pi}(1-|a|)}{(1+|a|)}e^{-\frac{1-|a|}{1+|a|}\xi}H(\xi), \quad (3.8)$$

where $H(\xi)$ is the Heaviside function.

We note that in either of the cases the spectrum contains nontrivial impulse at the origin. This prevents from direct use of Bedrosian's Theorem ([1]) in deducing mono- or dual mono-components of the form $\rho(t)e^{i\theta_a(t)}$ or $\rho(t)e^{i\phi_a(t)}$ with non-constant $\rho \geq 0$. We will come back to this in §6.

We present an example here.

Example. Taking $a = 1/2$ in Theorem 3.1, we have

$$e^{i\theta_a(t)} = c(t) + is(t),$$

where

$$c(t) = \frac{5\cos t - 4}{5 - 4\cos t}, \quad s(t) = \frac{3\sin t}{5 - 4\cos t}.$$

The theorem asserts that the function $c + is$ is a unimodular circular mono-component. As verification, now we prove this through a direct computation.

We will show

(i) $c^2(t) + s^2(t) = 1$;
(ii) $s'(t) = h(t)c(t), \quad h(t) \geq 0$;
(iii) $\widetilde{\mathcal{H}}\,(c(t)) = s(t)$.

Under (i)-(iii), we may write

$$c(t) = \cos\theta(t), \quad s(t) = \sin\theta(t), \quad \theta(t) = \int_0^t h(u)du, \quad \widetilde{\mathcal{H}}\,(\cos\theta(t)) = \sin\theta(t).$$

Now,

$$\begin{aligned} c^2(t) + s^2(t) &= \frac{(5\cos t - 4)^2 + (3\sin t)^2}{(5 - 4\cos t)^2} \\ &= \frac{25\cos^2 t + 16 - 40\cos t + 9\sin^2 t}{25 + 16\cos^2 t - 40\cos t}. \end{aligned}$$

On replacing 16 by $16\sin^2 t + 16\cos^2 t$ in the numerator, we have $c^2(t) + s^2(t) = 1$. This proves (i).

Next, we have

$$s'(t) = \left(\frac{3\sin t}{5 - 4\cos t}\right)' = \frac{15\cos t - 12}{(5 - 4\cos t)^2}.$$

Therefore

$$\frac{s'(t)}{c(t)} = \frac{3}{5 - 4\cos t} > 0.$$

To show (iii), we point out that

$$c(t) = -\frac{1}{2} + \frac{3}{4}\cos t + \ldots + \frac{3}{2^{n+1}}\cos nt + \ldots,$$
$$s(t) = \frac{3}{4}\sin t + \ldots + \frac{3}{2^{n+1}}\sin nt + \ldots.$$

The circular Hilbert transform may be applied term by term to the series $c(t)$, and we then obtain (iii). The term by term operation is justified by the L^2-boundedness of the circular Hilbert transform (Riesz Theorem) and by the convergence in the L^2-sense of the series representing function c .

From (i)-(iii), we know that, the phase

$$\theta(t) = \arctan\frac{\widetilde{\mathcal{H}}c(t)}{c(t)} = \arctan\frac{3\sin t}{5\cos t - 4},$$

satisfies that

$$\theta'(t) = h(t), \quad \cos\theta(t) = c(t), \quad \widetilde{\mathcal{H}}\cos\theta(t) = \widetilde{\mathcal{H}}c(t) = s(t) = \sin\theta(t).$$

Therefore, $e^{i\theta(t)}$ is a circular mono-component.

4. Boundary Values of Functions in Hardy Spaces

We first give some observations on boundary values of holomorphic functions.

If f belongs to $L^p(\mathbb{R}), 1 < p < \infty$, then the Cauchy integral

$$F(z) = \frac{1}{2\pi}\int_{\mathbb{R}}\frac{f(t)}{t - z}dt \tag{4.1}$$

is a holomorphic function in the upper-half complex plane, and Plemelj's formula holds:

$$\lim_{y\to 0+} F(x + iy) = \frac{1}{2}f(x) + i\frac{1}{2}Hf(x), \quad a.e.$$

The so called analytic signal, up to a multiple constant, is actually the boundary value of the Cauchy integral of the associated real signal. Distributionally,

$$\left(\frac{1}{(\cdot) - (x + iy)}\right)\hat{}\,(\xi) = \chi_{(0,\infty)}(\xi)e^{ix\xi}e^{-y|\xi|} = \frac{1}{2}(1 + \mathrm{sgn}\xi)e^{ix\xi}e^{-y|\xi|},$$

where $\chi_{(0,\infty)}$ denotes the characteristic function of the set $(0, \infty)$. A generalized Paseval's formula with (4.1) gives

$$F(z) = \frac{1}{2}\left[\frac{1}{2\pi}\int_{\mathbb{R}} e^{ix\xi}e^{-y|\xi|}\hat{f}(\xi)d\xi + i\frac{1}{2\pi}\int_{\mathbb{R}} e^{ix\xi}e^{-y|\xi|}(-i\mathrm{sgn}\xi)\hat{f}(\xi)d\xi\right].$$

Together with Plemelj's formula, this last representation in inverse Fourier transform shows that the spectrum of the Cauchy integral lies on the right-half of the real axis, and the Fourier multiplier of Hilbert transformation is $-i\text{sgn}\xi$.

The boundary value of an arbitrary holomorphic function in the upper-half complex plane does not necessarily have such property. The Cauchy integrals of functions in L^p spaces are actually functions in the Hardy H^p spaces. H^p spaces are large enough to fit our purpose. We introduce them as follows,

For $0 < p < \infty$, define the Hardy spaces

$$H^p(\mathbb{D}) = \{f \ : \ f \in H(\mathbb{D}),\ \|f\|_p = \sup_{0<r<1} \{\frac{1}{2\pi} \int_{\partial \mathbb{D}} |f(re^{it})|^p dt\}^{1/p} < \infty\},$$

and

$$H^p(\mathbb{C}^+) = \{f \ : \ f \in H(\mathbb{C}^+),\ \|f\|_p = \sup_{0<y<\infty} \{\int_R |f(t+iy)|^p dt\}^{1/p} < \infty\}.$$

For $p = \infty$, define

$$H^\infty(\mathbb{D}) = \{f \ : \ f \in H(\mathbb{D}),\ \|f\|_\infty = \sup_{z \in \mathbb{D}} |f(z)| < \infty\},$$

and

$$H^\infty(\mathbb{C}^+) = \{f \ : \ f \in H(\mathbb{C}^+), \|f\|_\infty = \sup_{w \in \mathbb{C}^+} |f(w)| < \infty\}.$$

For $p \geq 1, H^p$ are Banach spaces. For $p < 1, H^p$ are complete metric spaces under the metric

$$d(f, g) = \|f - g\|_p^p.$$

Since we are to study Hilbert transforms in our function spaces, we restrict ourselves to merely study $H^p(\mathbf{S}), 1 \leq p \leq \infty$.

In [15] we explore connections between eigenfunctions of Hilbert transformation and functions in Hardy H^p spaces. The following result is proved (Th. 3.2 and 4.3., [15]).

Theorem 4.1. *The function $f(t) = \rho(t)(c(t)+is(t))$, with $\rho \geq 0$ and $\rho \in L^p(\mathbf{S}), 1 \leq p \leq \infty, c^2 + s^2 = 1$, is the boundary value of a function in $H^p(\mathbf{S})$ if and only if $H_{\mathbf{S}}(\rho c) = \rho s$ modulo constants.*

We remark that $H_{\mathbf{S}}(\rho c) = \rho s$ is equivalent to $Hf = -if$. Note that for the frequently used case $\mathbf{S} = \mathbb{C}^+$ and $p = \infty$ the Hilbert transformation takes the distribution sense. Its formulation is naturally related to boundary values of functions in $H^\infty(\mathbb{R})$.

Denote by $\mathbf{D}$ the space of infinitely differentiable functions with compact support on the line, and $\mathbf{D}'$ the space of continuous linear functionals on $\mathbf{D}$, viz. the space of distributions.

Definition 4.2. Let T be a distribution and u a harmonic function in the upper-half complex plane. If

$$< T, \phi >= \lim_{y \to 0+} \int_{-\infty}^{\infty} u(x, y)\phi(x)dx, \quad \phi \in \mathbf{D},$$

then u is said to be a *harmonic representation of* T.

Obviously, a distribution may have more than one harmonic representations. The following result is known (see, for instance, [10] or [11]).

Theorem 4.3. *Let T be a distribution and U one of its harmonic representations. Let V be any harmonic conjugate of U, then V is a harmonic representation of some distribution, S.*

Definition 4.4. Any distribution S in Theorem 4.3 is called a *Hilbert transform* of T.

For a chosen harmonic representation of T its harmonic conjugates are not unique. As consequence the above defined Hilbert transform is unique only up to an additive constant. The relation $H^2 = -I$ now is changed to $H^2 = -I + [c]$, where $[c]$ denotes the class of constants.

Theorem 4.1 allows to obtain a large variety of unimodular H- and circular H-eigenfunctions through elementary conformal mappings.

We will be based on the following conformal mappings.

(i) The Cayley transform $\mathbb{C}^+ \to \mathbb{D}$ defined by

$$z = \kappa(w) = \frac{i - w}{i + w}.$$

The mapping $\kappa : \mathbb{C}^+ \to \mathbb{D}$ is univalent and onto.

(ii) The mappings $\mathbb{C}^+ \to \mathbb{D}$

$$\epsilon_L(z) = e^{i\frac{\pi z}{L}}, \quad L > 0.$$

They are onto but not univalent. They are periodic, satisfying $\epsilon_L(z + 2L) = \epsilon_L(z)$. Denote by $[\epsilon]$ the class of such mappings.

(iii) The Möbius transforms $\mathbb{D} \to \mathbb{D}$

$$\tau_a(z) = \frac{-\overline{a}}{|a|}\frac{z - a}{1 - \overline{a}z}, \quad a \in \mathbb{D}.$$

The conformal mappings are univalent and onto. We denote by $[\tau]$ the class of Möbius transforms.

(iv) The mappings $\mathbb{C}^+ \to \mathbb{C}^+$

$$\mu_{a,b,c,d}(z) = \frac{az + b}{cz + d}, \quad a, b, c, d \text{ real numbers, and } ad - bc > 0.$$

The conformal mappings are univalent and onto. We denote the class of such mappings by $[\mu]$.

(v) The mappings $\mathbb{C}^+ \to \mathbb{D}$

$$\nu_a(z) = \frac{|a^2+1|}{a^2+1}\frac{w-a}{w-\overline{a}}, \quad a \in \mathbb{C}^+ \setminus \{i\}.$$

The conformal mappings are univalent and onto. We denote the class of the mappings by $[\nu]$. Clearly, $\kappa \in [\nu]$.

(vi) Denote by $[f]$ and $[F]$ the classes of inner functions in $\mathbb{D}$ and $\mathbb{C}^+$, respectively. Inner functions are H^∞ functions with constant modulus 1 on the boundary.

(vii) Denote by $[b]$ and $[B]$ the classes of Blaschke products in $\mathbb{D}$ and $\mathbb{C}^+$, respectively.

We can construct functions in $[F]$ from the above listed elementary ones. Some examples are given in the following theorem.

Theorem 4.5. *We have*

(i) $[\tau] \circ [\epsilon] \subset [b] \circ [\epsilon] \subset [f] \circ [\epsilon] \subset [F]$.
(ii) $[\nu] = \kappa \circ [\mu] = [\tau] \circ \kappa \subset [b] \circ \kappa = [B] \subset [f] \circ \kappa = [F]$.
(iii) $[\tau] \circ [\epsilon] \circ [\mu] \subset [b] \circ [\epsilon] \circ [\mu] \subset [f] \circ [\epsilon] \circ [\mu] \subset [F]$.
(iv) *Products of functions in the classes in (i)-(iii) are functions in* $[F]$.

We draw the following remarks to the theorem.

Remark 4.6. The class $[b] \circ [\epsilon]$ in the assertion (i) consists of the unimodular mono-components on $\mathbb{R}$ of the form

$$F_1(t) = e^{i\sum \theta_{a_k}(\frac{\pi t}{L_k})}, \quad a_k \in \mathbb{D},\ L_k > 0,$$

where we take the convention that $\theta_{a_k}(t+2\pi) = \theta_{a_k}(t) + 2\pi$. The smallest class in (i) is $[\tau] \circ \epsilon_\pi$ that is the periodic version of Theorem 3.1.

Remark 4.7. The functions in the class $[\nu]$ of the assertion (ii) is the Cayley transform version of Theorem 3.2. They are the atomic cases of the class $[B]$. The class $[B]$ together with a factor of linear phase is studied in Picinbono [13].

Remark 4.8. The unimodular mono-components of the form $e^{i\mu(s)} = e^{i\frac{as+b}{cs+d}}$ are not periodic but with infinitely many oscillations. They are not Blaschke products, but belong to the class of "chirp " signals.

Remark 4.9. We may construct complicated mono-components based on the product rule specified in the assertion (iv). For instance, by multiplying the basic mono-component signals in (i), (ii) and (iii) we obtain, as long as convergent, the unimodular mono-component

$$e^{i\theta_0}\prod_{k=1}^{\infty}\frac{z_k - z}{z - \overline{z}_k}\exp(i\sum_{k=1}^{\infty}\theta_{w_k}(\frac{a_k z + b_k}{c_k z + d_k})),$$

where θ_0 is a real constant, $z_k, k = 1, 2, ...$, are complex numbers in the upper-half complex plane, $w_k, k = 1, 2, ...$ are complex numbers in the unit disc, and for each k, the real numbersa_k, b_k, c_k, d_k satisfy $a_k d_k - b_k c_k > 0$.

Next we give some spectrum results with distributions. We first introduce

Definition 4.10. Let T be a distribution and $f(x+iy)$ an analytic function in $\mathbb{C}^+$ such that for any $\phi \in \mathbf{D}$

$$< T, \phi >= \lim_{y \to 0+} \int_{\mathbb{R}} f(x+iy)\phi(x)dx,$$

then we say that T is an *upper-Hardy distribution*, and $f(x+iy)$ is an analytic representation of T. In such a case we may write $T = T^+$.

Let T be the tempered distribution represented by the boundary value of a function in $H^p(\mathbb{C}^+)$. From Definition 4.10, $T = T^+$ is an upper-Hardy distribution. The following theorem asserts that $\hat{T^+}$, the Fourier transform of T^+, has positive spectrum in the sense specified in the following theorem.

Theorem 4.11. *T^+ is the tempered upper-Hardy distribution represented by the boundary value of a function in $H^p(\mathbb{C}^+), 1 \le p \le \infty$, if and only if $\hat{T^+} \subset [0, \infty)$, that is,*

$$< \hat{T^+}, \phi >= 0, \text{ for all } \phi \in \mathbf{D} \text{ such that } \text{supp}\phi \subset (-\infty, 0].$$

The proof of the "only if" part is contained in [15], and that of "if" part is contained in [18]. Based on this theorem we generalized Bedrosian's Theorem to the L^p cases ([18]), as given in

Theorem 4.12. *Suppose $f \in L^p(\mathbb{R})$, $g \in L^q(\mathbb{R})$, $1/p + 1/q = 1/r$, $1 \le p, q, r \le \infty$. If $\text{supp}\hat{f} \subset [-\sigma, \sigma]$, $\text{supp}\ \hat{g} \subset \mathbb{R} \setminus (-\sigma - \delta, \sigma + \delta)$, where $\sigma, \delta > 0$, then*

$$H(fg) = fHg.$$

5. Boundary Values of Starlike Functions

In below, a connected and open set in the complex plane $\mathbb{C}$ is called a *domain*. A function f is said to be *univalent* if it takes different values at different points. Our definition for starlike domains, and therefore that for starlike functions, takes the narrow sense, that is, starlike with respect to the pole $z = 0$.

Definition 5.1. A domain Ω is said to be starlike if $0 \in \Omega$, and $tz \in \Omega, 0 < t < 1$, whenever $z \in \Omega$. A univalent and holomorphic function $f : \mathbb{D} \to f(\mathbb{D})$ is said to be *starlike* if $f(\mathbb{D})$ is starlike and $f(0) = 0$.

Closely related are *convex domains* and *convex functions*.

Definition 5.2. A domain Ω is said to be convex, if $0 \in \Omega$, and $tz_1 + (1-t)z_2 \in \Omega, 0 < t < 1$, whenever $z_1, z_2 \in \Omega$. A univalent and holomorphic function $f : \mathbb{D} \to f(\mathbb{D})$ is said to be *convex*, if $f(\mathbb{D})$ is convex and $f(0) = 0$.

Clearly, a convex domain is a starlike domain, and a convex function is a starlike function.

The Taylor expansion of a starlike function is of the form

$$g(z) = a_1 z + a_2 z^2 + \cdots + a_n z^n + \cdots, \quad |z| < 1. \tag{5.1}$$

We denote by S the class of univalent and holomorphic functions in $\mathbb{D}$ having the Taylor expansion

$$g(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots, \quad |z| < 1. \tag{5.2}$$

The totality of starlike functions in S is denoted by S^*, and the totality of convex functions in S is denoted by C. It may be shown that C is a proper subclass of S^*, and S^* is a proper subclass of S. We call functions in S^* *normalized starlike functions*; and those in C *normalized convex functions*. There has been a deep study with fruitful results on the classes C, S^* and S. Among literature on starlike functions we refer to [8], [6], [12], [4] and [7]. The most striking feature of the subtle analysis on the classes C, S^* and S would be its connections with Bieberbach conjecture (1916) whose final and celebrated proof was given by de Branges in 1984 ([8]). In this note we will explore some relations between starlike functions with our H-eigenfunction problem. We first introduce some concepts.

Definition 5.3. Let $\rho(t)$ and $\theta(t), 0 \le t \le 2\pi$, be absolutely continuous, $\rho \ge 0$, and

$$\int_0^{2\pi} \rho(t) e^{i\theta(t)} dt = 0. \tag{5.3}$$

With these properties, a function $f(t) = \rho(t)e^{i\theta(t)}$ is called a circular H-atom, if f is a circular mono-component satisfying $\theta(2\pi) - \theta(0) = 2\pi$; and, a dual circular H-atom, if f is a dual circular mono-component satisfying $\theta(2\pi) - \theta(0) = -2\pi$.

As consequence of Theorem 2.2, the following result addresses the symmetry property between circular and dual circular H-atoms.

Theorem 5.4. *$\rho(t)e^{i\theta(t)}$ is a circular H-atom if and only if $\rho(t)e^{-i\theta(t)}$ is a dual circular H-atom.*

The following results are contained in [6] (§1, Ch.10). If $f(z)$ is holomorphic, and it univalently maps $\mathbb{D}$ into a simply connected region Q whose boundary is a bounded rectifiable closed Jordan curve, then f continuously extends to $\bar{\mathbb{D}}$ such that on $\partial\mathbb{D}$ it is absolutely continuous with

$$\frac{df(e^{it})}{dt} = ie^{it} f'(e^{it}), \ a.e.,$$

where $f'(e^{it})$ is the non-tangential boundary value of $f'(z)$ in $\mathbb{D}$. If, moreover, $f(t) = \rho(t)e^{i\theta(t)}$ is the boundary value of a starlike function, then both $\rho(t)$ and $\theta(t)$ are absolutely continuous.

For practical reasons we only concern such ideal starlike functions. The importance of starlike functions lies on the following Theorem.

Theorem 5.5. *$\rho(t)e^{i\theta(t)}, 0 \le t \le 2\pi$, is a circular H-atom if and only if it is the boundary value $f(e^{it})$ of a starlike function $f(z)$ whose boundary is a bounded rectifiable closed Jordan curve.*

The proof is based on Argument Principle of univalent functions and the result for boundary values of functions in the $H^\infty(\mathbb{D})$ space ([15]). For details see [16].

In complex analysis the normalized starlike functions with respect to the pole ∞ have the Taylor expansion

$$f(z) = z + b_0 + \frac{b_1}{z} + \frac{b_2}{z^2} + \cdots \tag{5.4}$$

The existence of the first two entries, namely z and b_0, is mainly for a geometrically symmetric theory for this case. In particular, with the form (5.4), when $z = e^{it}$ goes along the unit circle in the anticlockwise direction, then $f(e^{it})$ goes along the boundary of $f(\mathbb{D})$ anticlockwise as well. For the theory of dual mono-component we, however, adopt the following definition that is analytically symmetric, and works well with Hilbert transform.

Definition 5.6. A function $f(z)$ is said to be starlike with respect to the pole ∞ if $f(\frac{1}{z})$ is starlike (with respect to the pole zero).

With this definition we have the counterpart result for dual circular H-atoms.

Theorem 5.7. *$\rho(t)e^{i\theta(t)}, 0 \le t \le 2\pi$, is a dual circular H-atom if and only if $\rho(t)e^{i\theta(t)}, 0 \le t \le 2\pi$, is the boundary value $f(e^{it})$ of a starlike function $f(z)$ with respect to the pole ∞, whose boundary is a bounded rectifiable closed Jordan curve.*

Example. (The Circle Family) The simplest example would be the circle family. Any fractional-linear transformation

$$w = f(z) = \frac{az}{cz + d}$$

that maps $\mathbb{D}$ into a disc $f(\mathbb{D}) \ni 0$, $f(0) = 0$, with the positive orientation as t rotates from 0 to 2π under the parametrization $z = e^{it}$, will give rise to a circular H-atom. We now form this family in a systematic way using Möbius transform. The Möbius transform $\tau_a(z) = (z - a)/(1 - \bar{a}z)$ has the power series expansion

$$\tau_a(z) = -a + b_1 z + b_2 z + \cdots,$$

where $b_1 = 1 - |a|^2 > 0$. We construct

$$f_a(z) = \frac{1}{b_1}(\tau_a(z) + a) = \frac{z}{1 - \overline{a}z}. \tag{5.5}$$

This function is in the class C. It maps discs in $\mathbb{D}$ into discs. The images $f_a(\mathbb{D}_r)$, $\mathbb{D}_r = r\mathbb{D}, 0 < r < 1$, are discs not centered at $z = 0$. Indeed,

$$f_a(re^{it}) = \frac{r}{\sqrt{1 - 2r|a|\cos(t - t_a) + |a|^2 r^2}} e^{i(t - \arg(1 - r|a|e^{i(t - t_a)}))},$$

where $a = |a|e^{it_a}$. It follows from Theorem 5.5 that for every fixed $r : 0 < r < 1$, the function $f_a(re^{it})$ is a circular H-atom. The mapping can be extended to $r : 1 \le r < 1/|a|$, and the diameter of the disc $f(\mathbb{D})$ passing through 0 is divided by 0 into two parts with lengths, respectively, $\frac{r}{1-r|a|}$ and $\frac{r}{1+r|a|}$. So, the closer the number $r|a|$ to 1, the closer the pole zero to the boundary of the image circle.

One can similarly formulate the ellipse family and the Casimire curve family.

As consequence of the Argument Principle finite products of circular H-atoms are multi-valent functions. We have the following

Theorem 5.8. *Finite products of circular and dual circular H-atoms are respectively circular mono-components and dual circular mono-components.*

The established theory on the classes S, S^* and C provides a source of starlike functions with a great variety. The basic references are [8], [6], [12], [4] and [7]. The reference [7], in particular, provides many working examples. We briefly recall, without proof, some results in the literature that may have significant impacts to our study.

(i) It may be shown that if $f(D)$ is starlike, then $f(D_r)$ is starlike for all $r \in (0,1)$. In Example 2.1 on the circle family we assert this fact from the property of fractional-linear transformations. It, however, holds in general. This implies that when $z = re^{it}$ traces out the circle $|z| = r$ anticlockwise, then the complex number $f(z) = \rho e^{i\theta}$ must also traces out a complete circle anticlockwise . It follows that

$$\frac{\partial}{\partial t} \arg\{f(z)\} = \frac{\partial \theta}{\partial t} \geq 0.$$

This latter condition implies

$$Re\{\frac{zf'(z)}{f(z)}\} \geq 0, \quad z \in \mathbb{D}.$$

This turns to be a sufficient condition for starlike domains as well.

(ii) It may be shown that a function is convex in $\mathbb{D}$ if and only if $1 + z\frac{f''(z)}{f(z)}$ has a positive real part in $\mathbb{D}$. As consequence, $f(\mathbb{D}_r), 0 < r < 1$, is also convex. Based on this it may be shown that $f(z)$ is convex if and only if $F(z) = zf'(z)$ is starlike. Therefore, a convex function $f(z)$ has the formula

$$f(z) = \int_0^z \frac{F(\zeta)}{\zeta} d\zeta,$$

where $F(z)$ is a starlike function. The last relation also gives rise to a representation formula for all convex functions (see (iv) below).

(iii) If f and g are in class S^*, then their weighted product $f^\alpha g^\beta, \alpha + \beta = 1, 0 \leq \alpha, \beta \leq 1$, is in S^*.

If f and g are in the class C with the expansions

$$f(z) = \sum_{n=1}^{\infty} a_n z^n, \quad g(z) = \sum_{n=1}^{\infty} b_n z^n,$$

then their Hadamard product (also called Hadamard convolution)

$$(f * g)(z) = \sum_{n=1}^{\infty} a_a b_n z^n$$

is in C.

If f and g are in the class S^*, then the modified Hadamard product

$$(f \otimes g)(z) = \sum_{n=1}^{\infty} \frac{a_a b_n}{n} z^n$$

is in S^*.

(iv) If $P(z)$ is holomorphic with positive real part then there holds Herglotz's formula:

$$P(z) = \int_0^{2\pi} \frac{e^{it} + z}{e^{it} - z} d\alpha(t),$$

where $\alpha(t)$ is a non-decreasing function satisfying

$$\int_0^{2\pi} d\alpha(t) = 1, \quad \textit{and} \quad \alpha(t) = \frac{1}{2}[\alpha(t+0) + \alpha(t-0)]. \tag{5.6}$$

There is a one to one relationship between the functions $P(z)$ and $\alpha(t)$.

Based on Herglotz's formula one has the representation formula for starlike functions: A function f is starlike in $\mathbb{ID}$ if and only if

$$f(z) = z \ \exp\left(2 \int_0^{2\pi} \log \frac{1}{1 - e^{-it} z} d\alpha(t)\right),$$

where α is a non-decreasing function satisfying (5.6). Theoretically the formula provides all starlike functions with the pole zero.

(v) It is an interesting fact that if $f(z)$ is in S, then for small enough $r > 0$ the image $f(r\mathbb{ID})$ is starlike, and therefore $f(rz)$ is in S^*. One can show that there exists a positive number, $R_{ST} = \frac{e^{\pi/2}-1}{e^{\pi/2}+1} \approx 0.65579$, called *radius of starlike-ness*, such that whenever $r \leq R_{ST}$ the image $f(r\mathbb{ID})$ is starlike for all $f \in S$. The number R_{ST} is sharp in the sense that if $r > R_{ST}$, then there exists a function $f \in S$ such that $f(r\mathbb{ID})$ is not starlike. For the class S there is also a sharp constant, $R_{CV} = 2 - \sqrt{3} \approx 0.26 \cdots$, called *radius of convexity*, such that whenever $r \leq R_{CV}$ the set $f(r\mathbb{ID})$ is convex for all $f \in S$.

It is the identical relationship between circular H-atoms and certain starlike functions that motivates the definition of circular H-atoms (Theorem 5.5). There is no counterpart concepts on the line. Next we will induce mono-components and dual mono-components on the line based on those obtained on the circle.

Theorem 5.9. *Assume that* $\tilde{f}(t) = \rho(t)e^{i\theta(t)}, 0 \leq t < 2\pi$, *where* $\rho \in L^p([0, 2\pi)), 1 \leq p \leq \infty$. *Then,*

(i) *for* $1 \leq p \leq \infty$, $\tilde{f}(t)$ *is a (dual) circular mono-component if and only if* $f(t) = \rho(t)e^{i\theta(t)}, -\infty < t < \infty$, *is a (dual) mono-component on the line, where* ρ *and* θ *are extended to satisfy* $\rho(t+2\pi) = \rho(t)$ *and* $\theta(t+2\pi) = \theta(t)+2\pi$.

(ii) *for* $1 \le p < \infty$, *the function*

$$\frac{1}{(s^2+1)^{1/p}}\rho(2\arctan s)$$

belongs to $L^p(\mathbb{R})$, *and, if* $\tilde{f}(t)$ *is a (dual) circular mono-component, then*

$$F(s) = \frac{1}{(s^2+1)^{1/p}}\rho(2\arctan s)e^{i\left(\theta(2\arctan s)+\frac{2}{p}\arccos(\frac{-s}{\sqrt{s^2+1}})-\frac{2\pi}{p}\right)}, \quad -\infty < s < \infty,$$

is a (dual) mono-component on the line.

(iii) *for* $p = \infty$, $\tilde{f}(t)$ *is a (dual) circular mono-component if and only if*

$$F(s) = \rho(2\arctan s)e^{i\theta(2\arctan s)}, \quad -\infty < s < \infty,$$

is a (dual) mono-component on the line.

The proof of (i) is based on the following Lemma.

Lemma 5.10. *Let* $\tilde{f} \in L^p([-\pi,\pi)), 1 \le p \le \infty$, *and* f *be the* 2π*-periodic extension of* $\tilde{f}$ *to the real line. Then* Hf *is* 2π*-periodic, and, restricted in* $[-\pi,\pi)$, $Hf = \tilde{H}\tilde{f}$, *where* Hf *is defined by*

$$Hf(t) = \lim_{\epsilon\to 0, N\to\infty} \frac{1}{\pi}\int_{\epsilon<|t-s|<(2N+1)\pi} \frac{f(s)}{t-s}ds.$$

Proof. It may be easily shown (also see [14], or [15], or [2])

$$\begin{aligned} Hf(t) &= \frac{1}{\pi}\lim_{\epsilon\to 0, N\to\infty}\int_{(-\pi,\pi)\cap\{|x-t|>\epsilon\}} \Big(\sum_{k=-N}^{N}\frac{1}{t-x-2k\pi}\Big)f(x)dx \\ &= \frac{1}{2\pi}\lim_{\epsilon\to 0}\int_{(-\pi,\pi)\cap\{|x-t|>\epsilon\}} \cot\left(\frac{t-x}{2}\right)f(x)dx \\ &= \tilde{H}\tilde{f}(t), \quad a.e. \end{aligned}$$

The proofs of (ii) and (iii) are based on Cayley transform and the relations between the two H^p spaces in the respective contexts. We suppress the easy proof. □

6. Solutions of the Eigenfunction Problem in Relation to Fourier Atoms

In this sections we deal with solutions of the eigenfunction problem of the forms $\rho(t)^{i\theta_a(t)}, 0 \le t \le 2\pi$, and $\rho(s)e^{i\phi_a(s)}, -\infty < s < \infty$. We construct some examples.

A signal f is said to be of λ-*low-frequency* , if $\hat{f}(\xi) = 0$ for $|\xi| > \lambda$; and of λ-*high-frequency*, if $\hat{f}(\xi) = 0$ for $|\xi| < \lambda$. Bedrosian's Theorem ([1], [17]) asserts that if f_l and f_h are in $L^2(\partial\mathbf{S})$, and f_l is of λ-low- and f_h of λ-high-frequency, then

$$H_{\mathbf{S}}(f_l f_h) = f_l H_{\mathbf{S}} f_h.$$

Note that in the above notion, when $\mathbf{S} = \mathbb{D}$, $\hat{f}(\xi)$ stands for the Fourier coefficients of f defined on integers. The discrete type Bedrosian's theorem is studied in [17]. A generalization of Bedrosian's Theorem to the L^p cases is given in [18].

Initially we wish to characterize $\rho(t)$ and $\theta(t)$ such that $\rho(t)e^{i\theta(t)}$ is a (circular) H-eigenfunction. In practice, the amplitude is expected to be of lower frequencies. In view of Bedrosian's theorem, we would expect

$$H_{\mathbf{S}}(\rho(t)\cos\theta(t)) = \rho(t)H_{\mathbf{S}}\cos\theta(t).$$

In such case, $\rho(t)e^{i\theta(t)}$ is a (circular) H-eigenfunction if and only if $e^{i\theta(t)}$ is a (circular) H-eigenfunction. This idea initiates the study of the unimodular case, that is $\rho \equiv 1$, of which a large class of unimodular (circular) mono-components is discovered ([17], [14], [15] and [2]). This, in particular, contains the Fourier atoms $e^{i\theta_a(t)}$ and their products. The question is: Having obtained the (circular) mono-components $e^{i\theta_a(t)}$, can we go back to identify functions ρ such that $\rho(t)e^{i\theta_a(t)}$ are (circular) mono-components, as already expected in view of Bedrosian's Theorem?

Bedrosian's Theorem may be used to construct non-unimodular mono-components of the form $\rho(t)e^{ikt}, k = \pm 1, \pm 2, ...$ The Fourier transform of e^{ikt} is the Dirac function $\delta(\cdot - k)$. Bedrosian's Theorem suggests that if ρ is non-negative and bandlimited, its Fourier transform has compact support contained in $[-|k|+\alpha, |k|-\alpha]$, where $|k| - \alpha > 0$, then $\rho(t)e^{ikt}$ may be a mono-component or dual mono-component, for $k > 0$ or $k < 0$, respectively.

Construction of such functions ρ is not entirely trivial. Without loss of generality, we can only deal with the case $k \geq 1$. First take any even and square-integrable function g_1 with compact support in $[-k + \alpha, k - \alpha]$, $1/2 > \alpha > 0$. Let g_0 be its inverse Fourier transform. Since g_1 is even, g_0 is real-valued. The Paley-Wiener Theorem (or a direct computation using inverse Fourier transform) implies that g_0 is the restriction of an entire function, $g_0(z)$, to the line. The entire function is in the Paley-Wiener $(k - \alpha)$-class satisfying the estimate

$$|g_0(z)| \leq Me^{(k-\alpha)|\mathrm{Im}z|}. \tag{6.1}$$

Note that $g_0(z)e^{i(k-\alpha)z}$ is a function in Hardy $H^2(\mathbb{C}^+)$ and $e^{i\alpha z}$ is in Hardy $H^\infty(\mathbb{C}^+)$, and hence their product is in $H^2(\mathbb{C}^+)$. Owing to the remark after the statement of Theorem 4.1 we have

$$\begin{aligned} H(g_0(t)e^{ikt}) &= H\left([g_0(t)e^{i(k-\alpha)t}][e^{i\alpha t}]\right) \\ &= -i[g_0(t)e^{i(k-\alpha)t}][e^{i\alpha t}] \\ &= -ig_0(t)e^{ikt}, \end{aligned}$$

showing that $g_0(t)e^{ikt}$ is an H-eigenfunction. For the constant M in (6.1) we also have

$$H(Me^{ikt}) = -iMe^{ikt}.$$

With $G(t) = g_0(t) + M > 0$ the signal $G(t)e^{ikt}$ is a mono-component.

Note that the above argument implies that $H(G(t)e^{ikt}) = G(t)H(e^{ikt})$ that is a Bedrosian type result but not under the standard conditions $f_l, f_h \in L^2(\mathbb{R})$.

Example. Let $k \geq 1$. Take $g_1 = \chi_{[-\delta,\delta]}, \delta < k - 1/2$, where $\chi_{[-\delta,\delta]}$ is the characteristic function of the interval $[-\delta, \delta]$. Then g_0, being inverse Fourier transform of g_1, is the sinc function $\frac{1}{\pi}\frac{\sin \delta t}{t}$ bounded by δ/π. Therefore,

$$\left(\frac{1}{\pi}\frac{\sin \delta t}{t} + \frac{\delta}{\pi}\right) e^{ikt}, \quad \delta < k - 1/2,$$

is a mono-component; and, owing to Theorem 2.2,

$$\left(\frac{1}{\pi}\frac{\sin \delta t}{t} + \frac{\delta}{\pi}\right) e^{-ikt}, \quad \delta < k - 1/2,$$

is a dual mono-component.

Bedrosian's Theorem, however, cannot be directly used to produce non-unimodular H-eigenfunctions of the form $\rho(t)e^{i\theta_a(t)}$. In fact, the results (3.7) and (3.8) show that for each of the two cases the spectrum contains nontrivial impulses at the origin. There, however, do exist non-unimodular functions $\rho \geq 0$ that make $\rho(t)e^{i\theta_a(t)}$ to be (circular) mono-components. This shows that the spectrum condition in the standard Bedrosian's Theorem is not a necessary one. We first deal with the circular case. Let τ_a be the Möbius transform defined in (3.2). Set

$$h_1(z) = \frac{1}{2}(\tau_a(z) + \overline{\tau_a(z^*)}),$$

where $z^* = 1/\overline{z}$ is the symmetric point of z with respect to the unit circle. Obviously, $h(z)$ is real-valued on the circle. Since $|\tau_a(z)|$ is dominated by 1, the function $h(z) = h_1(z) + 3$ has positive boundary values on the line, denoted by $\rho(t)$, between 1 and 5. The meromorphic function h has a sole pole of order one at a, so $G(z) = h(z)\tau_a(z)$ has no pole and is bounded and holomorphic in $\mathbb{D}$. We have $G(e^{it}) = \rho(t)e^{i\theta_a(t)}$. By invoking Theorem 4.1 for $H^\infty(\mathbb{D})$, we conclude that $\rho(t)e^{i\theta_a(t)}$ is a circular mono-component.

Simple computation gives

Example.

$$\begin{aligned} G(e^{it}) = \rho(t)e^{i\theta_a(t)} &= \left(2\frac{\cos t - 2|a|\cos t_a + |a|^2\cos(t-2t_a)}{1 - 2|a|\cos(t-t_a) + |a|^2} + 3\right) e^{\pm i\theta_a(t)}, \\ & 0 \leq t \leq 2\pi, \end{aligned} \tag{6.2}$$

where $a = |a|e^{it_a}, t_a \in \mathbb{R}$, is a circular or dual circular mono-component for the sign $\pm$ taking $+$ or $-$, respectively.

Example. Using the general result of §5, (i), Theorem 5.9 (also see [14] and [2]), the 2π-periodic extension of the above function,with $\theta_a(t + 2\pi) = \theta_a(t) + 2\pi$,

$$\begin{aligned} G(e^{it}) = \rho(t)e^{i\theta_a(t)} &= \left(2\frac{\cos t - 2|a|\cos t_a + |a|^2\cos(t-2t_a)}{1 - 2|a|\cos(t-t_a) + |a|^2} + 3\right) e^{\pm i\theta_a(t)}, \\ & -\infty \leq t \leq \infty, \end{aligned} \tag{6.3}$$

is a mono- or dual mono-component.

The non-periodic case on the real line may be deduced accordingly ((iii) Theorem 5.9)). The non-linear Fourier atoms $e^{i\theta_a(t)}$ on the circle under Cayley transform $\kappa : \mathbb{C}^+ \to \mathbb{D}$,

$$z = \kappa(w) = \frac{i-w}{i+w}, \quad e^{it} = \frac{i-s}{i+s} \tag{6.4}$$

is transformed to $e^{i\phi_a(s)}$ on the line, where

$$\phi_a(s) = \theta_a(2\arctan s).$$

Under the setting the function $e^{i\phi_a(s)}$ is the boundary value of the bounded and holomorphic function $\tau_a \circ \kappa$ that conformally and univalently maps the upper-half complex plane to the unit disc. Therefore $e^{i\phi_a(s)}$ is a mono-component. Similarly, take the function G defined right before Example (6.2), and set $\rho_1(s)e^{i\phi_a(s)} = (G \circ \kappa)(s)$. The obtained function is the boundary value of a bounded and holomorphic function in $\mathbb{C}^+$ and therefore

Example. In such a manner, with function ρ defined in (6.2), the function

$$\rho(2\arctan s)e^{\pm i\theta_a(2\arctan s)}, -\infty < s < \infty,$$

is a mono- or dual mono-component on the line.

Alternatively we can directly construct the same function with explicit representation using a similar method as in the circular case. Let $b \in \mathbb{C}^+$ such that $\kappa(b) = a$. Then $(\tau_a \circ \kappa)(b) = 0$. Apart from a unimodular multiple constant, the mapping $\tau_a \circ \kappa$, with boundary value $e^{i\phi_a(s)}$, is equal to $\kappa_b(z) = \frac{z-b}{z-\overline{b}}$. Set

$$h(z) = \frac{1}{2}(\kappa_b(z) + \overline{\kappa_b(\overline{z})})$$

which is meromorphic with a sole pole of order one at b, with bounded and real-valued boundary values on the real line. We define

$$G(z) = (h(z) + 3)\kappa_b(z).$$

It is a bounded and holomorphic function in $\mathbb{C}^+$.

Example. Apart from a unimodular multiple constant $G(z)$ has the boundary value

$$\left(2\frac{(s-s_b)^2 - h_b^2}{(s-s_b)^2 + h_b^2} + 3\right) e^{i\phi_a(s)},$$

where $b = s_b + ih_b$. The latter is a mono-component; and

$$\left(2\frac{(s-s_b)^2 - h_b^2}{(s-s_b)^2 + h_b^2} + 3\right) e^{-i\phi_a(s)}$$

is a dual mono-component.

References

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

[2] Q-H. Chen, L-Q. Li and T. Qian, Two families of unit analytic signals with nonlinear phase, accepted to appear in Physica D.

[3] Q-H. Chen, L-Q. Li and T. Qian, *Stability of frames generalized by non-linear Fourier atoms*, International Journal of Waveletss, Multiresolution and Information Processing, Vol. **3**, No.4 (2005), 1-12.

[4] P.L. Duren, *Univalent Functions,* Sprnger-Verlag, 1983.

[5] J. B. Garnett, *Bounded Analytic Functions*, Academic Press, 1987.

[6] G. M. Goluzin, *Geometric Theorey of Functions of a Complex Variable,* 2nd ed., Izdat. "Nauka": Moscow, 1966; English transl. Amer. Math. Soc., 1969.

[7] A. W. Goodman, *Univalent Functions,* Vol I, II, Mariner Publishing Co.,Tampa Florida, 1983.

[8] S. Gong, *The Bieberbach Conjecture*, AMS/IP, Studies in Advanced Mathematics, Volume 12, 1999.

[9] N. E. Huang et al, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, Proc. R. Soc. Lndon, A(1998)454, 903-995.

[10] B.H. Li, *On distributions with parameter and their analytic representations*, Chinese Math. Ann., 1981, 2(4), 399-405.

[11] B.H. Li and L.K. Guo, *Riesz transformations of distributions and a generalized Hardy space*, Approx. Theory and its Appl., 1988, 5(4), 1-17.

[12] Ch. Pommerrenke, *Univalent Functions,* Vanderhoeck and Puprecht; Göttingen, 1975.

[13] B. Picinbono, *On instantaneous amplitude and phase of signals*, IEEE Transactions on Signal Processing, vol, **45**, No. 3, March, 1997, 552-560.

[14] T. Qian, *Analytic Signals and Harmonic Measures*, J. Math. Anal. Appl., 314 (2006) 526-536.

[15] T. Qian, *Characterization of Boundary Values of Functions in Hardy Spaces With Application in Signal Analysis*, Journal of Integral Equations and Applications, Volume **17** Issue 2 (Summer 2005), 159-198.

[16] T. Qian, *Mono-components for decomposition of signals,* Math. Meth. Appl. Sci. 2006; **29**:1187-1198.

[17] T. Qian, Q-H. Chen, and L-Q. Li, *Analytic unit quadrature signals with nonlinear phase*, Physica D 203 (2005) 80-87.

[18] T. Qian, Y. Xu, D. Yan, L.-X. Yan and B. Yu, *Bedrosian theorem for L^p functions,* preprint.

Tao Qian
Faculty of Science and Technology
University of Macau,
Av. Padre Tomhs Pereira S.J.,
Taipa, Macau
e-mail: `fsttq@umac.mo`

Wavelet Analysis and Applications
Tao Qian, Mang I. Vai and Xu Yuesheng, Eds.
Applied and Numerical Harmonic Analysis, 321–362

Signal-Adaptive Aeroelastic Flight Data Analysis with HHT

Martin J. Brenner, Sunil L. Kukreja and Richard J. Prazenica

Abstract. This paper investigates the utility of the Hilbert-Huang transform for the analysis of aeroelastic flight data. The recently-developed Hilbert-Huang algorithm addresses the limitations of the classical Hilbert transform through a process known as empirical mode decomposition. Using this approach, the data is filtered into a series of intrinsic mode functions, each of which admits a well-behaved Hilbert transform. In this manner, the Hilbert-Huang algorithm affords time-frequency analysis of a large class of signals. The purpose of this paper is to demonstrate the potential applications of the Hilbert-Huang algorithm for the analysis of aeroelastic systems. Applications for correlations between system input and output, and amongst output sensors, are discussed to characterize the time-varying amplitude and frequency correlations present in the various components of multiple data channels. Examples are given using aeroelastic flight test data from the F/A-18 Active Aeroelastic Wing aircraft and Aerostructures Test Wing.

Mathematics Subject Classification (2000). Primary 93B15; Secondary 93B30.

Keywords. aeroelasticity, adaptive signal decomposition, Hilbert-Huang, empirical mode decomposition, time-frequency analysis.

1. Introduction

The Hilbert transform is a classical tool that has been used in the structural dynamics community as an indicator of nonlinearity. It has also been used to estimate nonlinear damping and stiffness functions for single degree-of-freedom systems.

The Hilbert transform expresses a signal as a harmonic with time-dependent frequency and amplitude. In this respect, it is an ideal tool for the analysis of nonstationary data. Unfortunately, the Hilbert transform has several shortcomings that limit its usefulness in practice. Most notably, the Hilbert transform computes a single instantaneous frequency for a signal at each instant in time. Therefore, when applied to a multi-component signal (i.e., a signal from a system with multiple modes), the Hilbert transform computes an instantaneous frequency that corresponds to a weighted average of the component frequencies. Such an instantaneous frequency does not provide any information as to the values of the individual component frequencies [4]. A further limitation is that the Hilbert transform yields grossly distorted estimates of the frequency when applied to signals with nonzero mean and signals which have more extrema than zero crossings.

In order to address these shortcomings, an Empirical Mode Decomposition (EMD) was developed by Huang et al. [8, 9] as a means of decomposing a signal into a series of components known as Intrinsic Mode Functions (IMFs). These IMFs are computed based on local characteristics of the signal and can be viewed as an adaptive, data-dependent basis. The IMFs form a complete, nearly orthogonal set of basis functions. Most importantly, each IMF contains only a single frequency component at any instant in time and therefore admits a well-behaved Hilbert transform. Taken collectively, the Hilbert spectra of the IMFs yield complete time-frequency information about the original signal. This approach, which has been termed the Hilbert-Huang Transform (HHT), makes it possible to apply the Hilbert transform to an extremely general class of functions and signals.

Although a relatively new tool, the Hilbert-Huang transform algorithm has received considerable attention in a number of engineering disciplines. This HHT algorithm has been applied in the analysis of scientific data [8, 9, 10], structural system identification [27, 28, 29], and mechanical system fault detection [11, 15, 26, 30]. A recent adoption into the image processing field [13, 14, 16, 31], the two-dimensional EMD is an adaptive image decomposition without the limitations from filter kernels or cost functions. The IMFs are interpreted as spatial frequency subbands with varying center frequency and bandwidth along the image. The EMD is a truly empirical method, not based on the Fourier frequency approach but related to the locations of extrema points and zero crossings. Based on this, the concept of "empiquency," used for time or space and short for "empirical mode frequency," was adopted to describe signal (image) oscillations based on the reciprocal distance between two consecutive extrema points. High concentrations of extrema points have high empiquency with sparse areas having low empiquency. Hence, applications for time-frequency-space signal processing are feasible.

A problem with the very versatile and most commonly used Morlet wavelet in dynamics data analysis is its leakage generated by the limited length of the basic wavelet function, which makes the quantitative definition of the energy-frequency-time distribution difficult. The interpretation of the wavelet can also be counterintuitive. For example, definition of a local event in any frequency range requires analysis in the high-frequency range, for the higher the frequency the

more localized the basic wavelet. A local event in the low-frequency range requires an extended period of time to discern it. Such interpretation can be difficult if possible at all. Another problem with wavelet analysis is its nonadaptive nature. Wavelet basis functions are predefined, whether of one type, or a multi-wavelet basis, or a dictionary of wavelets is selected. Data analysis is then constrained to these bases. Since the Morlet wavelet is Fourier based, it also suffers the many shortcomings of Fourier spectral analysis. A truly adaptive basis is a necessary requirement for nonstationary and nonlinear time series analysis and should be based on and derived from the data.

In this paper, results are obtained from the new approach. With the HHT, the intrinsic mode functions yield instantaneous frequencies as functions of time that give sharp identification of embedded structures. The main conceptual innovation in this approach is the introduction of the instantaneous frequencies for complicated data sets, which eliminate the need of spurious harmonics to represent nonlinear and nonstationary signals. This paper looks at the effect of enhancements like local/on-line versions of the algorithm [21]. To date, HHT analysis has only been performed on individual signals without regard to correlation with other data channels, or system inputs-to-outputs. Application for correlations between system signals are introduced to characterize the time-varying amplitude and frequency modulations present in the various components of multiple data channels including input and distributed sensors. In these respects, this paper attempts to elucidate the way EMD behaves in the analysis of F/A-18 Active Aeroelastic Wing (AAW) aircraft [18], aeroelastic and aeroservoelastic flight test data as well as Aerostructures Test Wing [12] and pitch-plunge simulation data.

2. Empirical Mode Decomposition

The classical Hilbert transform is, in principle, an effective tool for time-frequency analysis. Unfortunately, in practice, it can only be applied to an extremely restricted class of signals. In order to extend the utility of the Hilbert transform to more general signals, Huang et al. [8, 9] developed the empirical mode decomposition (EMD) as a means of preprocessing data before applying the Hilbert transform. The EMD procedure decomposes the original signal into a set of intrinsic mode functions (IMFs), each of which admits a well-behaved Hilbert transform. The Hilbert transform is then applied to each individual IMF, yielding an instantaneous frequency and amplitude for each IMF. Therefore, this procedure, termed the Hilbert-Huang transform (HHT), enables the Hilbert transform to be applied to multi-component signals.

There are two criteria that each IMF must satisy in order to be amenable to the Hilbert transform. Namely, each IMF must have zero mean and the number of local extrema and zero crossings in each IMF can differ by no more than one. The first step in the EMD procedure is to connect all the local maxima of the original signal using a cubic spline. Similarly, the local minima are also connected with a cubic spline. The two splines define the envelope, and the mean of this envelope

is then calculated and subtracted from the original signal. The resulting signal is then tested to see if it satisfies the criteria for an IMF. If it does not, the sifting process is repeated until a suitable first IMF, $c_1(t)$, is obtained. The sifting process is then applied to the residual signal $x(t) - c_1(t)$ to obtain the next IMF. This process is repeated until all that is left is a final residual, $r(t)$, which represents the trend in the data and is not an IMF. Therefore, as shown in Eq. 1, the EMD procedure yields a decomposition of the original signal in terms of n IMFs and the residual, $r(t)$.

$$x(t) = \sum_{j=1}^{n} c_j(t) + r(t) \tag{1}$$

The number of IMFs obtained is dependent on the original signal.

The EMD procedure serves to generate IMFs that are amenable to the Hilbert transform. In particular, the manner in which the EMD is performed guarantees that each IMF will only possess a single harmonic at any instant in time. Currently, the EMD procedure is ad hoc in the sense that there is no rigorous mathematical theory behind it. Recent attempts are being made to formalize EMD, most notably the work of those who explored the properties of B-splines and their use in the EMD process. Despite the lack of a firm theoretical foundation, several mathematical properties of the IMFs are well understood [20]. In particular, it is clear from Eq. 1 that the IMFs, along with the residual, form a complete basis for the original signal $x(t)$. In addition, Huang et al. [8] demonstrated that the IMFs are nearly orthogonal with an orthogonality index (OI_{mn} defined in Eq. 2) for the IMFs. This definition seems to be global but actually only applies locally. Adjacent IMFs could have data with the same frequency but at different times.

$$OI_{mn} = \sum_{t=0}^{t=T} \sum_{j=1}^{m} \sum_{k(\neq j)=1}^{n} \frac{c_j(t)c_k(t)}{x(t)^2} \tag{2}$$

The EMD does not yield a unique basis for the original signal since there are countless sets of suitable IMFs that can be generated from a given signal. However, the various IMF sets from the different sifting criteria are all equally valid representations of the data provided their orthogonality indices are sufficiently small [8, 10]. The IMFs depend on the stoppage criterion, maximum number of siftings, intermittance criteria, the end point boundary conditions, and use of curvature- or extrema-based sifting [8, 9, 10]. The uniqueness problem can only be meaningful if all these parameters are fixed a priori. The problem is how to optimize the sifting procedure to produce the "best" IMF set. These questions are difficult to answer theoretically. In Huang et. al. [10], a confidence limit is defined for the first time without invoking the ergodic assumption. This provides a stable range of stopping criteria for the EMD-sifting operation, thereby making the HHT method more definitive. Discussion later points out that the EMD sifting process also acts in such a manner as to obtain IMFs that correspond to approximate bandpass filtering of the original signal.

Once the EMD procedure has been used to generate a set of IMFs, the Hilbert transform can be applied to each individual IMF. In this manner, an instantaneous frequency and amplitude is computed for each function. A common method for displaying the Hilbert spectrum is to generate a two-dimensional plot with time and frequency axes. The amplitude is then plotted as a color spectrogram in the time-frequency plane. By plotting the Hilbert spectra of all the IMFs together one obtains complete time-frequency information about the original signal.

To demonstrate the importance of the EMD in obtaining meaningful Hilbert spectra for general signals, consider the following signals.

$$\begin{aligned} x_1(t) &= \sin(2\pi t) + \sin(8\pi t) + \sin(32\pi t) \\ x_2(t) &= \sin(4\pi t) + t \end{aligned}$$

The signal x_1 is a combination of harmonics with frequencies of $\{1, 4, 16\}$ Hz while the signal x_2 is composed of a ramp signal and a 2-Hz sine wave. First, consider the Hilbert spectra obtained by directly applying the Hilbert transform to each signal. Figure 1 displays the resulting Hilbert spectra for both signals. The Hilbert spectrum of x_1 yields a fluctuating frequency with most of the energy concentrated at 7 Hz. This corresponds to the average of the three frequencies present in x_1 (each of which has the same amplitude) and gives no useful information about the original signal. This averaging is due to the fact that the Hilbert transform computes a single instantaneous frequency for the signal. The Hilbert spectrum of x_2 illustrates the effect of a nonzero mean. In this case, the estimated frequency is distorted and the 2-Hz harmonic is not identified at all.

Figure 2 displays the signal x_1 and its decomposition into four IMFs (imf1=c_1, imf2=c_2, etc.) and a residual. Clearly, each IMF has captured a different frequency component of the original signal. The fourth IMF is extremely small and results from boundary effects in the EMD process. The Hilbert spectra of the individual IMFs are plotted together in Fig. 2. Now, an effective Hilbert spectrum, or Hilbert-Huang spectrum, of x_1 has been obtained, with all three frequency components clearly identified. Similarly, Fig. 3 depicts two IMFs and a residual generated from the signal x_2. The 2-Hz frequency component has been sifted into the first IMF and the ramp component has been identified as the second IMF. The mean of the signal is 2, which has been separated out as the residual. The Hilbert spectrum clearly shows the 2-Hz component and estimates a low frequency for the ramp component. This occurs because the ramp is treated as part of a low frequency wave. The IMFs and Hilbert spectra in Figs. 2 and 3 illustrate that there are some minor boundary effects associated with the EMD process. Most importantly, these examples demonstrate that EMD makes it possible to apply the Hilbert transform to signals that otherwise do not admit well-behaved Hilbert transforms.

Finally, two more examples [21] illustrate automatic and adaptive (signal-dependent) time-variant filtering of general mixtures of signals. A signal composed of three components which significantly overlap in time and frequency is successfully decomposed in Fig. 4 for the sum of two sinusoidal frequency modulations and one Gaussian wave packet. Another example, accenting the nonharmonic nature

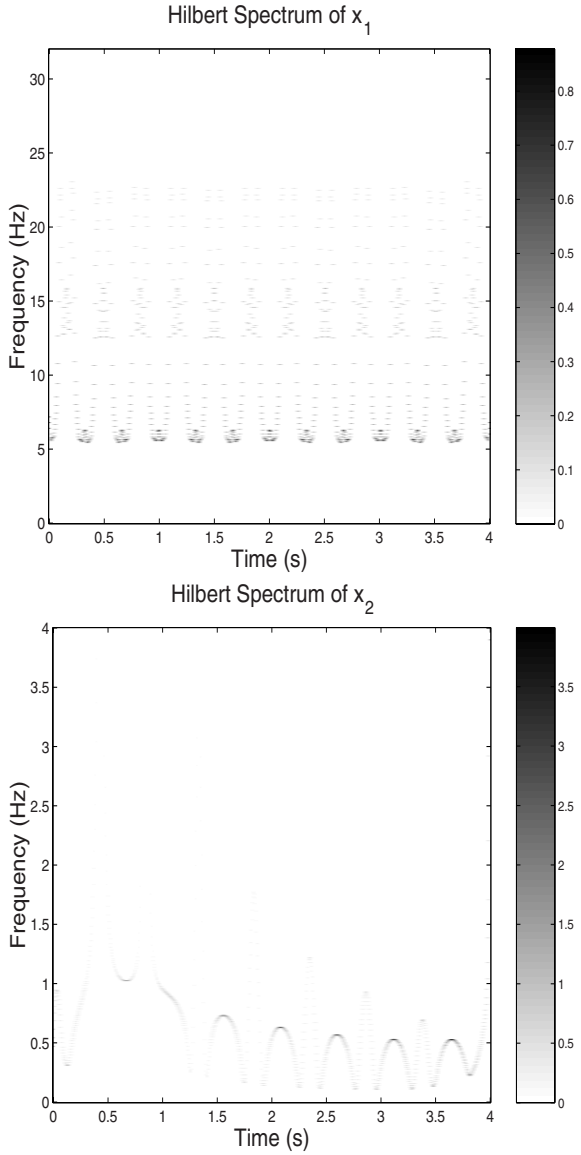

FIGURE 1. Time-frequency spectra from the Hilbert transforms of the signals x_1 and x_2.

of EMD, is given in Fig. 5. The analyzed signal (top) is the sum of three components, a sinusoid superimposed on two triangular waveforms with periods smaller and larger than the sinusoid. The decomposition performed by the EMD is given in

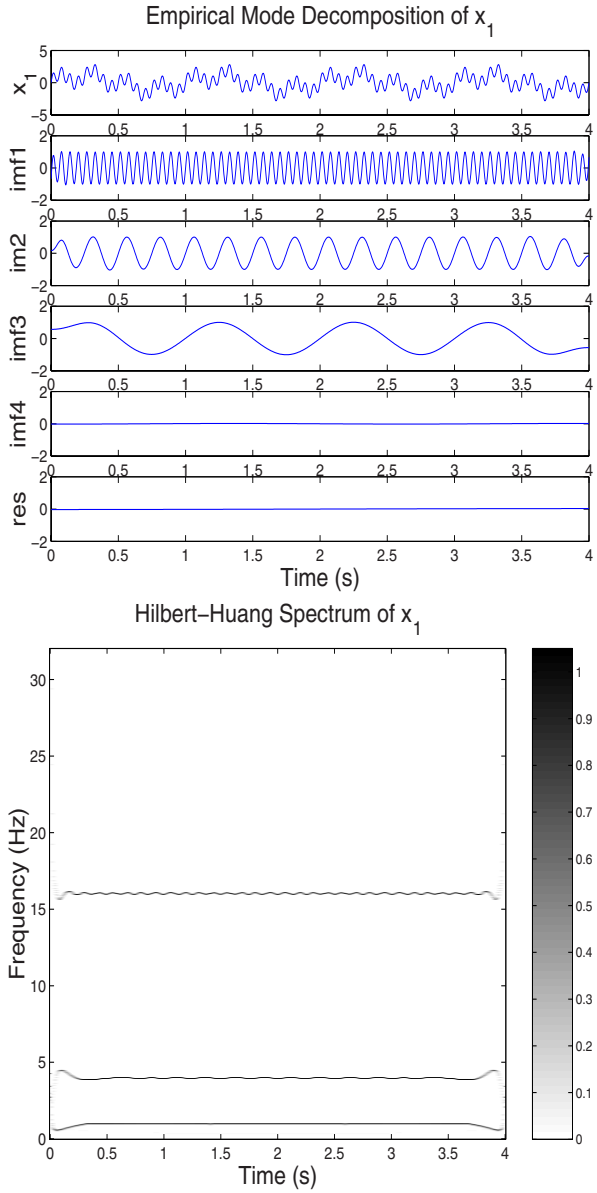

FIGURE 2. Empirical mode decomposition and the Hilbert-Huang spectrum of the signal x_1.

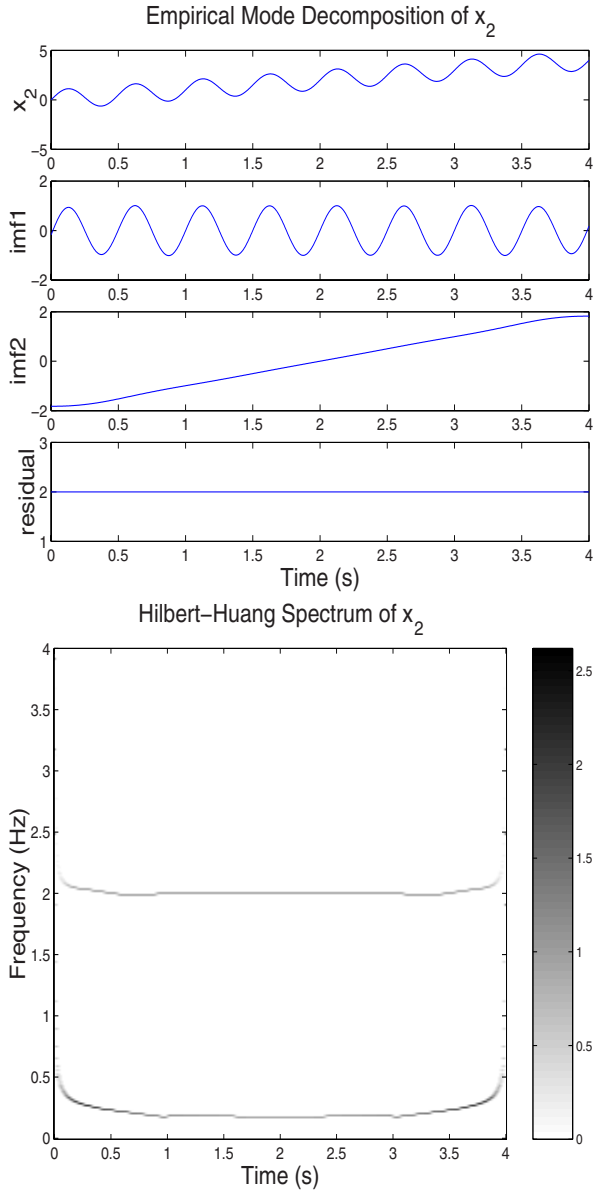

FIGURE 3. Empirical mode decomposition and the Hilbert-Huang spectrum of the signal x_2.

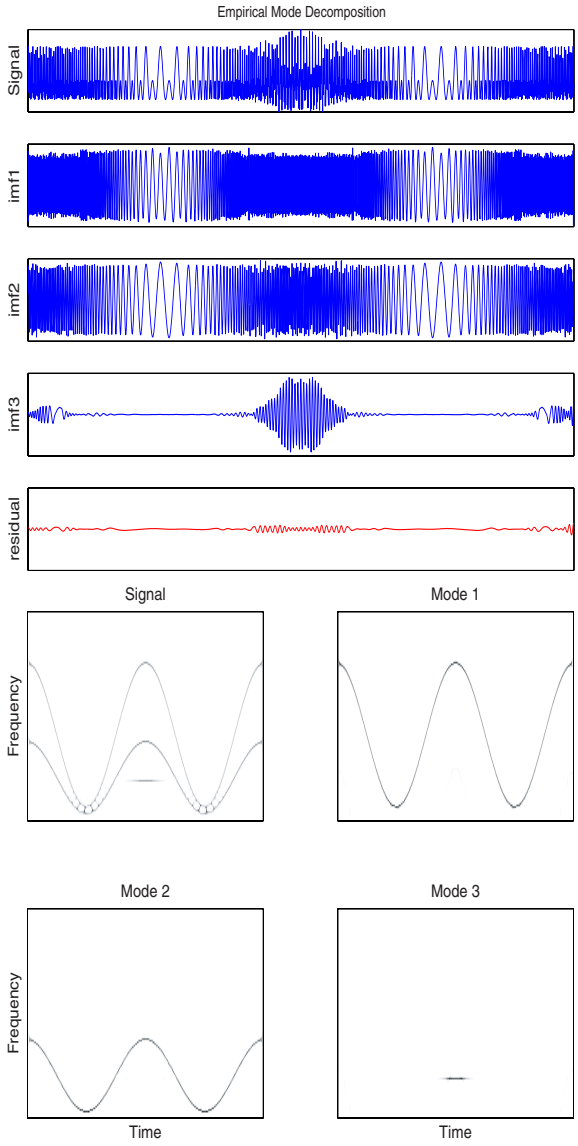

FIGURE 4. Empirical mode decomposition (EMD) and the Hilbert-Huang spectrum of the sum of two sinusoidal frequency modulations (FMs) and one Gaussian wave packet. Original signal (top), FM1, FM2, Gaussian wave packet, and residual.

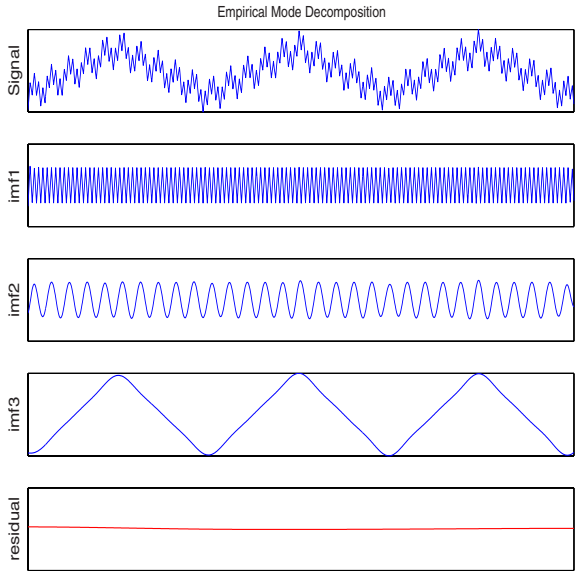

FIGURE 5. Empirical mode decomposition of the sum of one sinusoid with two triangular waveforms.
Original signal (top), high-frequency triangular, sinusoid, low-frequency triangular, and residual.

the three IMFs and the residual. In this case, both linear (sinusoid) and nonlinear (triangular) oscillations are effectively identified and separated. Harmonic analysis (Fourier, Morlet wavelets) would produce a less succinct and less descriptive decomposition.

2.1. Filtering Properties

The filtering properties of EMD have been studied in some detail [6, 7, 25]. The EMD process yields a data-dependent decomposition that focuses on local characteristics of the signal. In particular, EMD sifts out the highest-frequency component in the signal at any given time. Indeed, EMD has surprisingly been shown to behave as a dyadic bandpass filter when decomposing Gaussian white noise, much like a multiresolution wavelet decomposition. However, the EMD method as an equivalent dyadic filter bank is only in the sense of its global behavior over the entire time extent. In representing the time-frequency distribution, the Hilbert spectrum of each IMF is actually localized at any time. This is different from predetermined filtering such as with Morlet wavelet filtering.

2.2. Analytical Interpretation

The EMD is faced with the fundamental difficulty of not admitting an analytical definition, but of rather being defined by the algorithm itself, thus making

the analysis of its performance and limitations difficult. The need for rigorous mathematical foundation is imperative. This fundamental problem of the empirical mode decomposition has to be resolved since only with the intrinsic mode function can nonlinear distorted waveforms be resolved from nonlinear processes. There have been attempts to circumvent the mathematical difficulties in the EMD with some success by casting the IMFs in terms of B-splines [3], and applying them towards mechanical system fault-detection [15]. System identification of the IMFs as a multi-component system is suggestive in the light of multiresolution system identification procedures such as multiresolution singular value decompositions, Kalman filters, and subspace algorithms.

More significantly, characterization of IMFs as solutions to certain self-adjoint ordinary differential equations is demonstrated [22, 23]. Construction of envelopes which do not rely on the Hilbert transform is used directly to compute the coefficients of the differential equations. These equations are natural models for linear vibrational problems and provide further insight into both the EMD procedure and utilizing its IMF components to identify systems of differential equations naturally associated with the components. One of the uses of the EMD procedure is to study solutions to differential equations, and vibration analysis was a major motivation in the development of the Sturm-Liouville theory.

3. Hilbert Transform and Instantaneous Frequency

The Hilbert transform of a time-domain function or signal x is defined in Eq. 3,

$$y(t) = \mathrm{H}\{x(t)\} = \frac{1}{\pi}\ \mathrm{PV} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{3}$$

where PV denotes the Cauchy principal value, needed because the integrand is singular at $\tau = t$. The Hilbert transform can be viewed as the convolution of the original signal with $1/t$, emphasizing temporal locality of $x(t)$. Note that, unlike Fourier analysis, the Hilbert transform of a time-domain signal is another time-domain signal. The Hilbert transform is sometimes applied to frequency-domain signals using a similar expression as Eq. 3, but this paper will focus on the time-domain case. In practice, the Hilbert transform is usually calculated using the Fourier transform. Therefore, the fast Fourier transform algorithm can be employed for the efficient calculation of the Hilbert transform.

A signal, x, and its Hilbert transform, y, can be used to define a complex analytical signal as in Eq. 4.

$$z(t) \quad = a(t)\mathrm{e}^{i\theta(t)} \quad = x(t) + iy(t) \tag{4}$$

Therefore, the Hilbert transform pair $\{x(t), y(t)\}$ can be expressed as a harmonic function with time-varying amplitude $a(t)$ and time-varying phase angle $\theta(t)$.

$$\begin{aligned} a(t) &= \sqrt{x(t)^2 + y(t)^2} \\ \theta(t) &= \tan^{-1}\left(\frac{y(t)}{x(t)}\right) \end{aligned}$$

Given the time-dependent phase angle, the instantaneous frequency of the signal can be defined as [5]

$$\omega(t) = \frac{d\theta(t)}{dt}.$$

In the context of Eq. 3 with instantaneous frequency, the Hilbert transform of an IMF can be interpreted as giving the best fit with a sinusoidal function to the data weighted by $1/t$. Instantaneous frequency can be computed using the derivative definition or centralized finite difference. [5] In general, there are an infinite number of ways to express a signal as in Eq. 4, so there can also be an infinite number of instantaneous frequencies. The Hilbert-transform pair was proposed to uniquely define the amplitude and phase by building the complex analytic signal from the given signal with the original signal, $x(t)$, as the real part and the orthogonal transformed signal, $y(t)$, the imaginary part, out of phase with $x(t)$ by $\frac{\pi}{2}$. Now given the set of Hilbert-transformed IMFs and associated instantaneous frequencies $\omega_j(t)$ for each IMF component, the Hilbert spectrum, $H(\omega, t)$, is defined as the time-frequency distribution of the IMF amplitudes.

$$H(\omega_j, t_i) = a_j(t_i), \forall\, \omega_j(t_i) \quad \Longleftrightarrow \quad x(t) = \sum_{j=1}^{n} a_j(t) \mathrm{e}^{i \int \omega_j(t) dt} \tag{5}$$

Note that Fourier analysis yields the decomposition

$$x(t) = \sum_{j=0}^{\infty} a_j e^{i\omega_j t}$$

which is similar to the form of the Hilbert transform in Eq. 5. A key difference, however, is that the Fourier decomposition is in terms of harmonics with constant amplitudes and frequencies. In contrast, the Hilbert transform yields instantaneous amplitudes and frequencies. Therefore, in principle, the Hilbert transform is an ideal tool for the time-frequency analysis of a general class of signals, including nonstationary signals.

Another important distinction between Fourier analysis and the Hilbert transform is that the Fourier decomposition is in terms of multiple harmonics of constant amplitude and frequency, thereby producing artificial harmonics to maintain energy conservation for nonstationary and nonlinear data. In contrast, the Hilbert transform of a signal yields an expression in terms of a single harmonic with a time-varying frequency and amplitude. For this reason, the Hilbert transform is only suitable for the analysis of mono-component signals, or signals that are composed of a single frequency component at any instant in time. This is a considerable limitation as it implies that the Hilbert transform cannot be directly applied to signals that are composed of multiple harmonics. As was shown, the Hilbert transform fails to identify the individual frequencies and instead computes a single instantaneous frequency that corresponds to a weighted average of the component frequencies. The resulting instantaneous frequency is both physically invalid and erroneous [4] since a multi-component signal has more than one instantaneous frequency. An additional limitation of the Hilbert transform is that

it yields extremely distorted estimates of the instantaneous frequencies of signals with nonzero mean and signals that have more local extrema than zero crossings.

The EMD responds to the dilemma surrounding the applicability of instantaneous frequency. It decomposes a multi-component signal into its associated mono-components while not obscuring or obliterating the physical essentials of the signal and allows the traditional definition of instantaneous frequency to be complete by being applicable to signals of both mono- and multi-component. To follow the true frequency evolution within a multi-component signal, it is necessary to break down the components into individual and physically meaningful intrinsic parts. The adaptive and nonarbitrary decomposition using EMD produces an orthogonal set of intrinsic components each retaining the true physical characteristics of the original signal. The mono-components or intrinsic modes satisfy the conditions for a well-defined notion of instantaneous frequency. These conditions include symmetry, no dependence on predefined time scales, revelation of the nature of simultaneous amplitude and phase variation, and near-orthogonality.

Finally, since instantaneous frequency displays frequency variation with time, changes of dynamic states indicative of nonlinearity can be identified. For example, if the instantaneous frequency of a new mode is about half of the frequency of the old mode in a bifurcation, period doubling occurs. If the instantaneous frequency of the new mode is disproportionate with the old mode, quasi-periodic bifurcation occurs. Similarly, intermittence and chaotic motion can be determined. A dynamic state can be diagnosed simultaneously by observing the changes in time of the instantaneous frequency components and their corresponding energy. In summary, the concept of mode defined in relation to instantaneous frequency as a periodic-modal structure in the instantaneous time-frequency plane is found to be more appropriate than artificial sinusoidal harmonics in characterizing nonlinear responses [26]. Instantaneous frequency is a quantity critical for understanding nonstationary and nonlinear processes.

4. Local On-Line Decompositions

In the original EMD formulation, sifting iterations are applied to the predefined full length signal as long as there exists a locality at which the mean of the upper and lower envelopes is not considered sufficiently small enough. Excessive iterations on the entire signal to achieve a better local approximation contaminates other parts of the signal by overcompressing the amplitude and overdecomposing it by spreading out its components over adjacent intrinsic modes; i.e., overiteration leads to overdecomposition. The various stoppage criteria (to fulfill that the number of extrema and the number of zero crossings must differ at most by one, and that the mean between the upper and lower envelopes must be close to zero) are attempts to avoid the rigor of the symmetry of the envelopes without a mathematically rigorous definition for an adaptive basis. The hierarchical and nonlinear nature of the EMD algorithm will not provide that the EMD of segmented signals will be the segmentation of individual EMDs. Therefore, a variation referred to as local

EMD [21] introduces an intermediate step in the sifting process. Localities at which the error remains large are identified and isolated, and extra iterations are applied only to them. This is achieved by introducing a weighting function such that maximum weighting is on those connected segments above a threshold amplitude, with a soft decay to zero outside those supports, much like soft-thresholding is done in wavelet denoising.

Another option is based on the idea that sifting relies on interpolations between extrema, and thus only requires a finite number of them (five minima and five maxima in the case of the recommended cubic splines [3, 8]) for local interpolation. Extraction of a mode could therefore be moving blockwise instead of globally over the entire time span. This led to the development of the on-line EMD [21]. A prerequisite for the sliding window extraction of a mode is to apply the same number of sifting steps to all blocks in order to prevent possible discontinuities. Since this would require the knowledge of the entire signal, the number of sifting operations is proposed to be fixed a priori to a number less than 10 for effective application of the on-line version of the EMD algorithm operating in coordination with the local EMD described above. The leading edge of the window progresses when new data become available, whereas the trailing edge progresses by blocks when the stopping criterion is met on a block. Therefore, an IMF and corresponding residual are computed sequentially, then again applied to this residual, thus extracting the next mode with some delay.

An aileron command multisine input used on the F/A-18 AAW for aeroservoelastic response and flutter clearance is shown in Fig. 6 (top plots) using the standard EMD (left) and local/on-line version (right). The bandpass nature of IMFs [6, 7, 25] is reflected in the three standard IMF mean frequencies, $\{23.9, 16.3, 7.7\}$ Hz for each of IMFs $\{\#1, \#2, \#3\}$, respectively, and on-line corresponding IMF mean frequencies $\{23.6, 13.0, 7.2\}$ Hz. Immediately noticeable is the more efficient extraction of the signal components by the local/on-line algorithm, most evident by the second and third IMFs (imf2 and imf3) being more sparse than the corresponding standard IMFs. Besides the obvious advantage of an on-line algorithm for decomposing data, it has been found to clearly surpass the standard global algorithm in terms of computational burden, especially with long original data records. An added bonus is that it generally has better orthogonality properties among the IMFs, witnessed by an order of magnitude improvement in the orthogonality index defined in Eq. 2.

5. Local Analytic Signal Correlation

Since the IMFs allow permissible, meaningful, physically sensible, and unique interpretations of instantaneous frequency of general signals of interest, they fit into the class of asymptotic signals such that the time variation of the IMF amplitude and frequency may directly be recovered from the time variation of the amplitude and of the phase derivative of the associated analytic signal. An IMF after performing the Hilbert transform can be written as in Eq. 4. These complex components

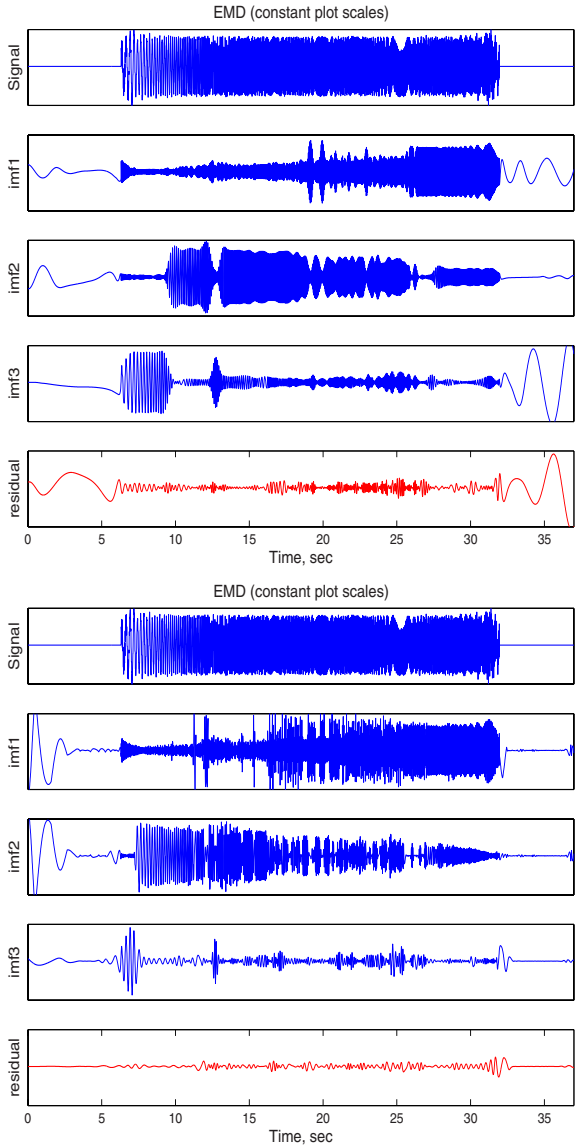

FIGURE 6. Standard (top) and local/on-line (bottom) empirical mode decomposition of an F/A-18 AAW multisine aileron command input, with original signal at top and residual at bottom.

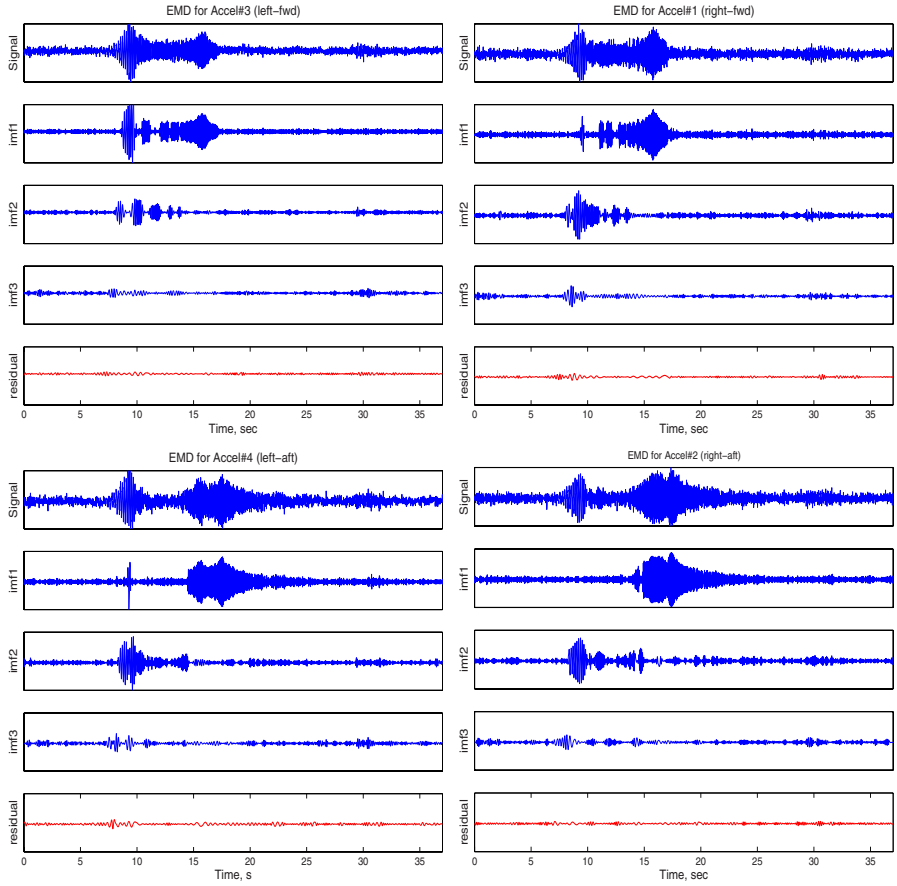

FIGURE 7. EMDs (constant vertical plot scales) of F/A-18-AAW aircraft left-fwd wingtip accel (top-left, Accel#3), right-fwd wingtip accel (top-right, Accel#1), left-aft wingtip accel (bottom-left, Accel#4), and right-aft wingtip accel (bottom-right, Accel#2), responses from the multisine symmetric aileron command shown in Fig. 6 (original signal at top and residual at bottom in each set).

are now used for analysis of analytic data correlations [1, 2] between input-output and amongst spatially distributed sensor outputs. Note that these analyses are all local in nature since there are no assumptions of stationarity, and differ from classical double-time expressions [1] by being instantaneous.

5.1. Local Correlation Coefficient

Correlations are made between transformed IMFs of various signals given the associated complex analytic signals

$$\begin{aligned} Z_x(t) = &\quad A_x(t)e^{i\phi_x(t)} &= x(t) + ix_H(t) \\ Z_y(t) = &\quad A_y(t)e^{i\phi_y(t)} &= y(t) + iy_H(t) \end{aligned}$$

by considering the *cross-analytic* signal defined in Eq. 6 by

$$\begin{aligned} Z_{xy}(t) &= Z_x^*(t)Z_y(t) = A_x(t)A_y(t)e^{i\phi_{xy}(t)} \\ &= A_x(t)A_y(t)e^{i[(\omega_y-\omega_x)t+(\phi_{0y}-\phi_{0x})]} \end{aligned} \tag{6}$$

which is not necessarily analytic but can be used to characterize the time variation of the phase difference between components *at similar frequencies* extracted from two simultaneously acquired signals. This Hilbert demodulation technique is only applicable for analytic components, in this case the transformed IMFs of the original signals. If the two components are mainly in-phase and a change occurs in the cross-analytic signal, the instantaneous frequencies of the two components may be drifting apart and this produces an instantaneous contribution to the phase difference from Eq. 6. A measure of the local correlation between components, in terms of simultaneous changes in instantaneous amplitude or frequency (phase) between analytic signals, is the Hilbert Local Correlation Coeffcient, $HLCC$.

$$HLCC(t) = \quad \frac{Re[Z_{xy}]}{|Z_{xy}|} \quad = \cos[\phi_{xy}(t)]$$

As a simple illustration of the utility of the $HLCC$, Fig. 7 shows the EMDs of wingtip accelerometer responses (accels) due to a symmetric aileron input, with the top plot in each set being the original signal, followed by EMDs 1-3 and the residual. The top two sets represent the forward (fwd) wingtip accels, and corresponding IMFs seem not to compare quite as well as the bottom two IMF sets of aft wingtip accels for the same input. The bottom IMF sets seem much more correlated with each other than the top two sets.

Table 1 lists the means, medians, and standard deviations of the $HLCC$ data over the time span of the maneuver. Note that $HLCC_{xy}^{mn}$ indicates correlation coefficients between Accel#m and Accel#n, using cross-analytic IMF function Z_{xy} correlating Accel#m's IMFx to Accel#n's IMFy. Results from the table indicate strongest Hilbert local correlation coefficients for Z_{xy} when $x = y$, as expected since this is a correlation between similar bandpass characteristics of IMFs from different responses due to the same input. For comparison, the standard statistical signal correlation coefficient between Accel#1 and #3 is $C^{13} = 0.8633$, and between Accel#2 and #4 is $C^{24} = 0.8332$ (correlating fwd-to-fwd accels and aft-to-aft accels, respectively, wingtip-to-wingtip). This is consistent with the top original signal plots in Fig. 7 in that the forward wingtip accels (top two, fwd) seem to correlate reasonably well with each other, as also between the bottom two (aft), but not top compared with the bottom. Evident in Table 1 is the common standard deviations (STD in all the correlations, near $STD = 0.7$). In the $HLCC$ values good commonality is found between the mean and medians comparing $HLCC^{13}$

and $HLCC^{24}$ in that the trend is similar amongst the cross-analytic IMF functions Z_{xy} where x and y correspond to respective IMF numbers in different accel EMDs. This is surprising, since in Fig. 7 the correlation in corresponding IMFs seems much worse between the two fwd accels (top plot IMFs) than between the two aft accels (bottom plot IMFs).

TABLE 1. $HLCC$ results from F/A-18 AAW wingtip accelerations.

Z_{xy}	Mean $HLCC^{13}$	Median $HLCC^{13}$	STD^{13}
Z_{11}	0.3283	0.5968	0.6801
Z_{12}	0.0917	0.1782	0.7171
Z_{13}	0.0098	0.0252	0.7113
Z_{21}	0.0953	0.1870	0.7148
Z_{22}	0.1844	0.3500	0.7048
Z_{23}	0.1187	0.2364	0.7136
Z_{31}	-0.0071	-0.0175	0.7075
Z_{32}	0.0845	0.1753	0.7097
Z_{33}	0.2403	0.4425	0.6866
Z_{11}	Mean $HLCC^{24}$	Median $HLCC^{24}$	STD^{24}
Z_{12}	0.0502	0.1080	0.7142
Z_{13}	0.0005	-0.0106	0.7057
Z_{21}	0.0492	0.0896	0.7095
Z_{22}	0.2197	0.4012	0.7023
Z_{23}	0.0692	0.1319	0.7052
Z_{31}	0.0032	0.0062	0.7056
Z_{32}	0.0996	0.2025	0.7183
Z_{33}	0.1553	0.2877	0.7027

Figure 8 shows the $HLCC$ functions between Accel#1 and Accel#3 ($HLCC_{xy}^{13}$) at 5-20 s. Decomposing and representing correlations like this allows a true time-localized instantaneous measure revealing subtle properties in the data and inconspicuous relations to other data sets using analytic components.

5.2. Instantaneous Transfer Function

In addition to investigating correlations between sensors, an instantaneous system transfer function is introduced with instantaneous magnitude and the $HLCC$ as a phase parameter between input-output analytic signals. This is viable in terms of the HHT inducing analytic properties to the IMFs to yield localized system input-output properties. Instantaneous transfer function (ITF), its instantaneous magnitude (IM), and its instantaneous phase (IP) are defined.

$$ITF(t) = \frac{Z_{xy}(t)}{Z_{xx}(t)}; \quad IM(t) = |ITF(t)|; \quad IP(t) = \cos^{-1}[HLCC(t)] = \phi_{xy}(t)$$

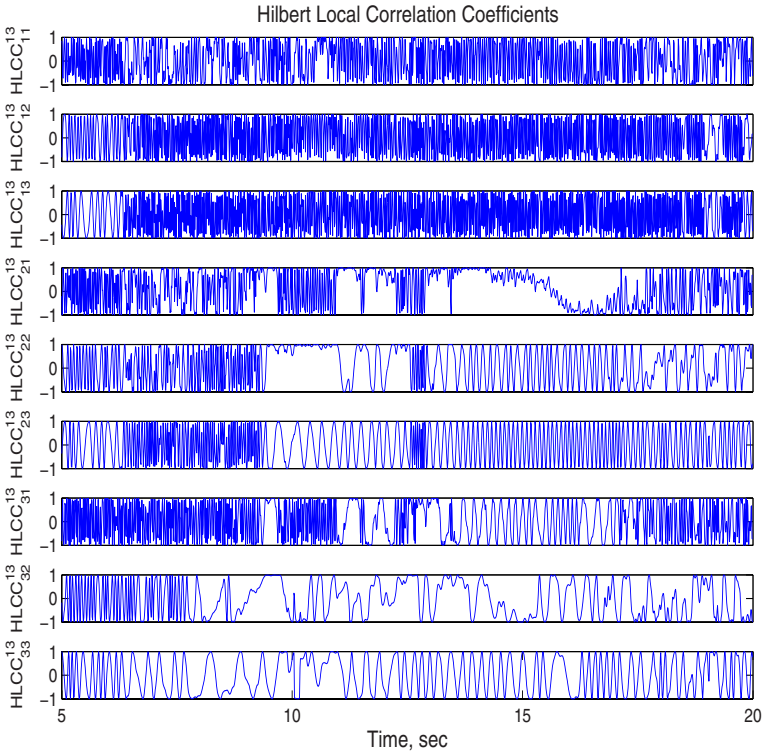

FIGURE 8. $HLCC^{13}_{xy}$ functions between IMFs of Accel#1 (top right, Fig. 7) and Accel#3 (top left, Fig. 7).

In Fig. 9 the top set of plots represent instantaneous magnitude (IM^{a1}_{xy}) and the bottom plots represent instantaneous phase (IP^{a1}_{xy}) from the instantaneous transfer function (ITF^{a1}_{xy}) of the aileron command input (Fig. 6) to Accel#1 (top right, Fig. 7) analytic IMFs (Hilbert-transformed i.e., $\{Z_x, Z_y\}$). Note that ITF^{mn}_{xy} indicates transfer functions between Input#m and Accel#n, using cross-analytic IMF function Z_{xy} correlating Input#m's IMFx to Accel#n's IMFy. In this sense, by interpreting the input-output Hilbert-transformed pairs of IMFs as a multi-component system of input-output signals, local stability measures are deemed to be feasible by tracking gain, phase, and instantaneous frequencies between each IMF pair. How these local IMF properties correspond to global system properties, given the analytic transformed IMFs, and aeroelastic and aeroservoelatic applications in stability and health monitoring, are currently being researched.

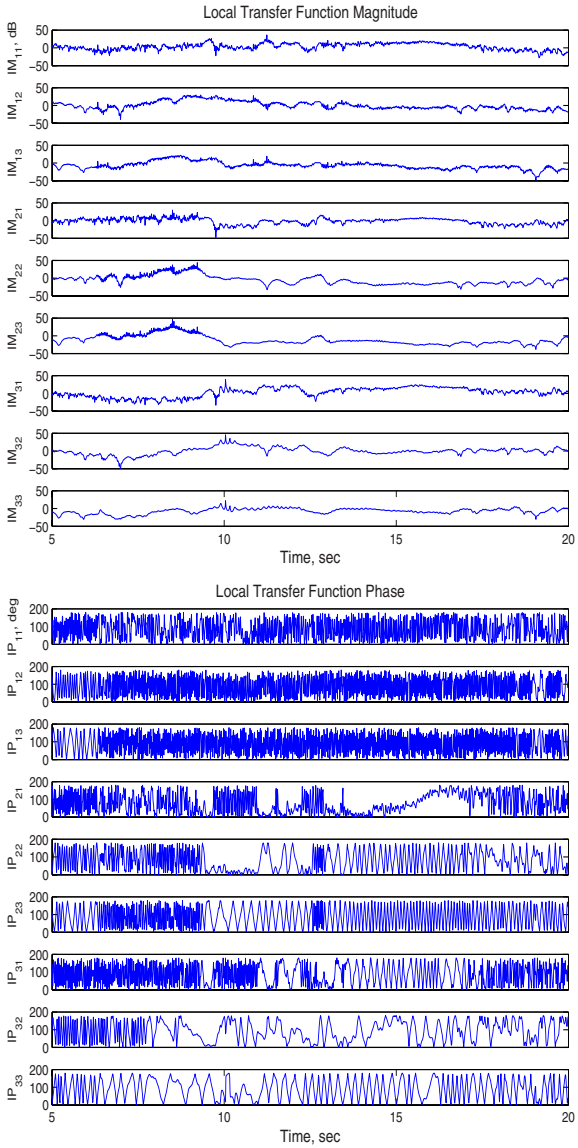

FIGURE 9. Instantaneous log-magnitude (IM_{xy}^{a1}, top plot set) and phase (IP_{xy}^{a1}, bottom plot set) of IMF transfer functions from F/A-18 AAW aileron input (Fig. 6) to Accel#1 (top right, Fig. 7).

5.3. Empirigrams and Empirical Local Correlation Coefficient

Given two EMDs from different signals, select a common number of IMFs to use for correlation (as three were selected in earlier examples for input and outputs). This is generally not difficult, especially for aeroelasticity data, since the higher-numbered IMFs approach a residual characteristic quickly. From the transformed IMFs, $\{Z_x(t), Z_y(t)\}$, define the corresponding set of two-dimensional Hilbert *empirigrams*, $\{H_x(\eta_x, t), H_y(\eta_y, t)\}$, where for each common $\eta = \eta_x = \eta_y$, the IMF number from the input EMD corresponds with the output EMD. Define in Eq. 7 the Hilbert *cross-empirigram*

$$H_{xy}(\eta, t) = H_x^*(\eta, t) H_y(\eta, t) \tag{7}$$

which correlates the respective IMFs from the two Hilbert empirigrams. Empirigrams relate to time-scale wavelet scalograms [5], which relate to time-frequency maps, since scales relate to frequencies in standard wavelet decompositions. Because of the bandpass nature of IMFs discussed previously, a similar construction emanates with the HHT. As with wavelet scalograms, the real part of the cross-empirigram (co-empirigram) gives the instantaneous contribution of each IMF to the correlation between two signals. An Empirical Local Correlation Coefficient, $ELCC(\eta, t)$, is then defined as

$$ELCC(\eta, t) = \frac{Re[H_{xy}(\eta,t)]}{|H_{xy}(\eta,t)|} = \frac{Re[H_{xy}(\eta, t)]}{|H_x(\eta, t)||H_y(\eta, t)|}$$

where $ELCC(\eta, t)$ (between ± 1, as the $HLCC$) gives the instantaneous contribution between corresponding IMFs from the two signals to the correlation coefficient. Figure 10 shows an imaged decomposition plot of the $ELCC$ between the input signal from Fig. 6 to Accel#1 in Fig. 7, but includes all nine IMFs from the EMDs of the input and output. The first row represents the contributions of the first input IMF to first output IMF, etc., up to the ninth IMF. In each IMF row there is much oscillation (higher frequency in the first and lower frequency to the ninth) of contributions from corresponding IMFs to the correlation. There are generally stronger correlations over longer time spans in the higher-numbered IMFs (lower frequencies), but the lower-numbered IMF correlations are less obvious due to the higher frequency content. There is a tendency to cycle from high-to-low-to-high [strong(positive)-to-none-to-strong(negative)] correlation very rapidly. A more detailed depiction in the zoomed-in bottom plot, between 10-11 s, demonstrates a rich interplay between correlation of mid-to-lower IMFs (higher frequencies) over the shorter time period. Yet another view is presented in Fig. 11, where the contours are split up discretely in three dimensions showing the heavy emphasis in the higher frequencies because the contours are more congested in each of the IMF levels. These representations highlight the areas of commonality and incongruity between corresponding input-output IMFs.

From the Hilbert spectrum $H(\omega, t)$ the energy spectrum $H^2(\omega, t)$ gives instantaneous energy,

$$IE(t) = \int_\omega H^2(\omega, t) d\omega$$

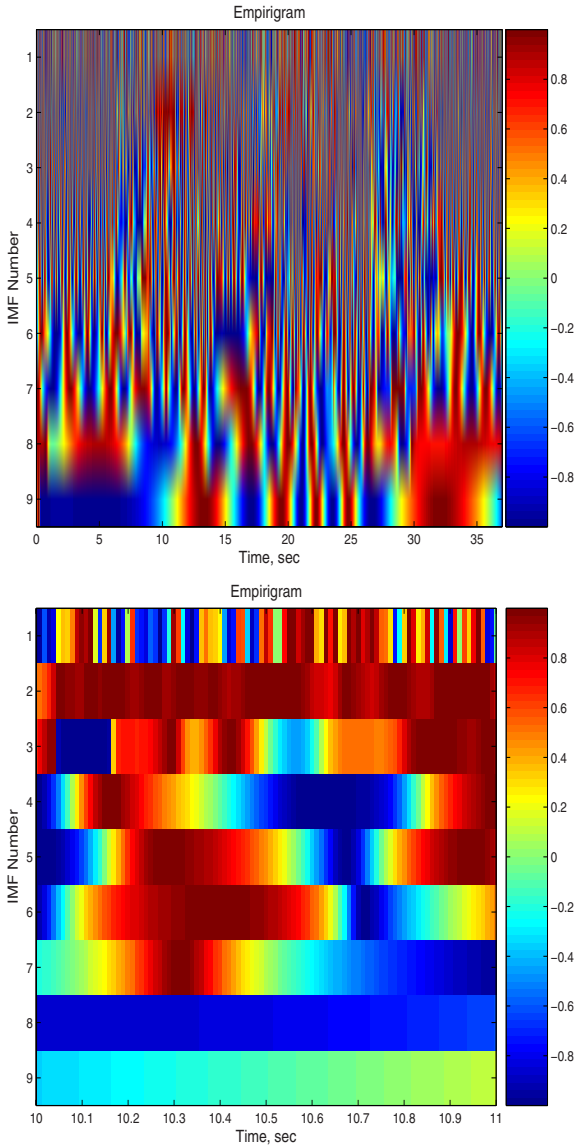

FIGURE 10. Empirical local correlation coefficient $ELCC$ plots using two different representations of the $ELCC$ between the input signal from Fig. 6 to Accel#1 in Fig. 7. Top plot is an intrinsic mode component-by-component depiction, starting from the first at the top to the ninth at the bottom, where individual intrinsic functions are clearly delineated. Bottom plot is zoomed-in portion between 10-11 s of the same $ELCC$.

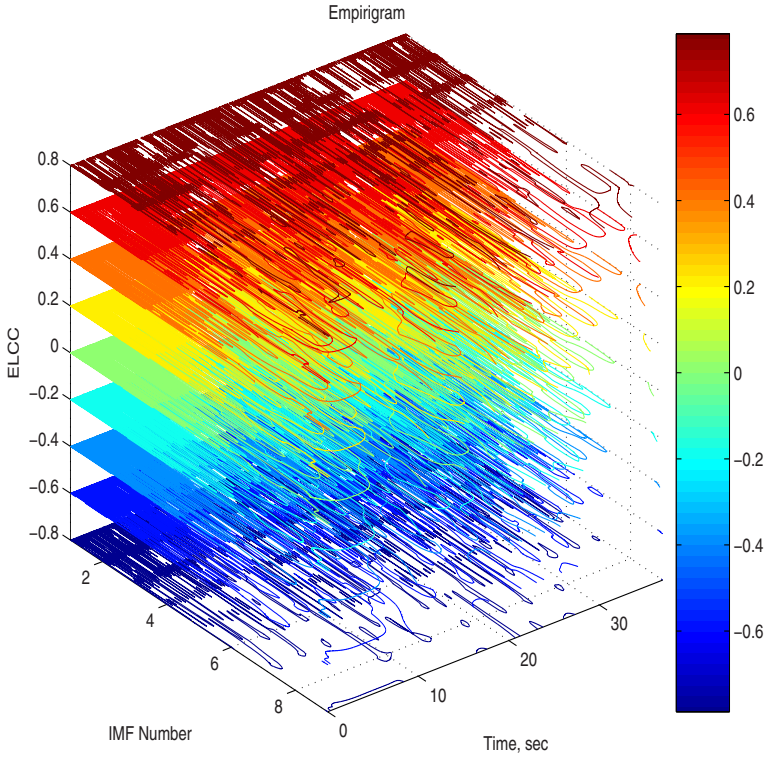

FIGURE 11. *ELCC* represented using a three-dimensional contour of the result in Fig. 10.

which is an indication of the energy fluctuations with time being weighted by the Hilbert spectrum localized energy over the entire set of IMFs. Corresponding to the signals in Fig. 7 are the instantaneous energy profiles of the output accelerometer responses in Fig. 12 (top set of four plots). The aft wingtip accels (bottom plots) are very similar (from the symmetric aileron input), indicating energy at $\{7-10, 13-20, 30\}$ s time locations. In this case, with the particular multisine input from Fig. 6 programmed over the $3-35$ Hz frequency range, these times correspond closely to the primary F/A-18 AAW modal frequencies in a corresponding frequency range, i.e., $\{6-9, 12-20, 30-35\}$ Hz, as will be shown with marginal Hilbert spectra. The forward accelerometer responses in the top plots are also very similar indicating modes near $\{6, 12-17, 25-32\}$ Hz. Instantaneous energy over a sensor suite is therefore a nonstationary indicator of time-varying energy distribution amongst the sensor array.

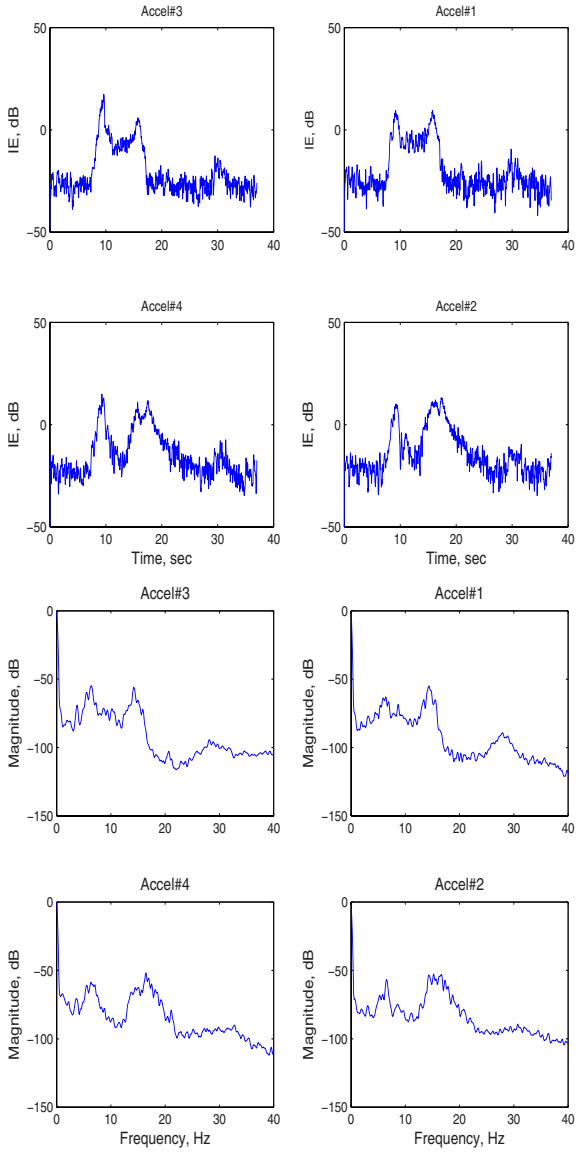

FIGURE 12. Instantaneous energy (top plot set) and marginal spectra (bottom plot set) of right-fwd wingtip accel (top right, Accel#1), right-aft wingtip accel (bottom right, Accel#2), left-fwd wingtip accel (top left, Accel#3), and left-aft wingtip accel (bottom left, Accel#4) responses, respectively in each set, from the multisine symmetric aileron command shown in Fig. 6.

6. Marginal Spectra

From the Hilbert spectrum $H(\omega, t)$, the marginal spectrum is also defined.

$$h(\omega) = \int_t H(\omega, t)dt$$

Corresponding to the signals in Fig. 7 are the marginal spectra of the output accelerometer responses in Fig. 12 (bottom set of four plots). The aft wingtip accels (bottom plots) are again similar showing modal response at $\{5-10, 12-20, 30-35\}$ Hz. The forward accelerometer responses in the top plots are also again very similar indicating modes near $\{6, 12-17, 25-32\}$ Hz.

The marginal spectrum measures total amplitude (or energy) contributions from each frequency value over the entire data record in a probabilistic sense. Frequency in either $H(\omega, t)$ or $h(\omega)$ is very different from Fourier spectral analysis. In the Fourier analysis the existence of energy at a frequency implies a wave component persisting through the data. Energy at the marginal Hilbert frequency, however, implies a higher likelihood (expected value over time) for such a wave to have appeared locally. As stated earlier, the Hilbert transform of an IMF gives the best fit with a sinusoidal function to the data weighted by $1/t$, which makes it instantaneous. In fact, the Hilbert spectrum is a weighted nonnormalized joint amplitude-frequency-time distribution in which the weight assigned to each time-frequency cell is the local amplitude. Consequently, frequency in the marginal spectrum indicates only the likelihood that an oscillation with such a frequency exists. Exact occurrence of frequency content is given in the full Hilbert spectrum. The Fourier spectrum is meaningless for nonstationary data, and there is little similarity between Hilbert and Fourier from previous studies [8, 9] for nonstationary data. Also, marginal cross-spectra between signals does not make sense since the time-dependence (causality) is lost and the frequencies are time-ignorant, so an input-output correlation at a certain frequency is meaningless.

The time-frequency Fourier spectrum, a spectrogram, suffers from the same restrictions over time due to windowing distortions. There is a lower bound on the local time-frequency resolution uncertainty product of the spectrogram due to the windowing operation. *This limitation is an inherent property of the spectrogram and is not a property of the signal or a fundamental limit* [17]. For many other time-frequency distributions, the local uncertainty product is less than that of the Fourier spectrogram and can be arbitrarily small. These results are contrary to the common notion that the uncertainty principle limits local quantities. Similar considerations apply to window-based filter bank (wavelet) methods. This limitation is due to the window and not to any inherent property of the signal.

This is a key point for application and understanding of HHT analysis, that windowing is not a factor, so standard time-frequency resolution limitations based on the uncertainty principle do not apply. As seen in the on-line analysis, an adaptive sliding window method has good performance simply requiring adequate data in the window to initiate a sifting process for satisfying the two IMF properties for analyticity.

7. Wing Dynamics Analysis

Dynamics between different sensors from the same input will now be investigated. This information will be used to guide the analysis of a group of wing accelerometer responses from a single input in terms of input-output correlation contribution for the F/A-18 AAW aircraft [18] and Aerostructures Test Wing (ATW) [12].

7.1. F/A-18 Active Aeroelastic Wing (AAW) Aircraft

Collective F/A-18 AAW aileron position, used as the multisine input, was obtained as the average of four position transducer measurements from the right and left ailerons. Outputs are twelve wing structural accelerometers located at the left (six accels) and correponding right (six accels) outer wings, all sampled at 400 sps, as designated in Table 2.

TABLE 2. F/A-18 AAW wing accelerometer nomenclature.

3. L-wt-fwd 3. L-wingtip forward	1. R-wt-fwd 1. R-wingtip forward
7. L-wof-fwd 7. L-wing outer-fold forward	5. R-wof-fwd 5. R-wing outer-fold forward
11. L-wif-fwd 11. L-wing inner-fold forward	9. R-wif-fwd 9. R-wing inner-fold forward
4. L-wt-aft 4. L-wingtip aft	2. R-wt-aft 2. R-wingtip aft
8. L-wof-aft 8. L-wing outer-fold aft	6. R-wof-aft 6. R-wing outer-fold aft
12. L-wif-aft 12. L-wing inner-fold aft	10. R-wif-aft 10. R-wing inner-fold aft

EMDs were calculated for the input and all outputs, with mean frequencies calculated for all the output IMFs compared to the single input IMFs. Table 3 is a summary of the results. Included are the first six IMFs averaged amongst the accels, percentage differences of these compared to the input IMFs, and standard deviations of output sensor IMFs (numbered 1-12). The accels correspond very well amongst each other in IMF frequencies (as frequency decreases, from left to right). The comparison to the input frequency is excellent even though input and output EMDs are performed independently. Orthogonality of the IMFs in each case was also very good. This shows that for this type of data, where responses are all from a common input (whether it is known or not), IMFs can be used for sensor-to-sensor correlation and input-output analysis.

Now the concept of the using the essentially-orthogonal analytic IMFs from the inputs and outputs is pursued to establish a multi-loop connotation of input IMFs to output IMFs for correlation and even stability properties. The input IMFs are interpreted as an orthogonal decomposition of the input(s), and the same for output IMFs for output(s). This can be generalized to multi-input-multi-output

TABLE 3. IMF frequencies from F/A-18 AAW input collective aileron position and outer wing response data.

Signal	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
0. Input	116.62	59.61	29.34	16.75	9.05	4.90
1.	99.05	56.94	27.90	12.56	6.96	3.62
2.	123.85	71.36	35.15	16.83	9.35	5.02
3.	114.47	66.05	33.78	15.06	7.95	5.32
4.	127.56	78.71	41.73	21.06	11.92	6.58
5.	105.83	58.52	32.06	17.17	9.34	5.47
6.	96.31	54.96	29.24	14.91	8.64	4.63
7.	115.93	58.77	30.88	16.59	8.66	4.32
8.	102.99	52.76	27.14	15.35	8.65	4.83
9.	112.19	60.61	31.99	17.78	9.90	5.17
10.	115.39	64.07	36.09	19.05	10.29	5.72
11.	124.91	66.06	32.99	18.32	10.53	5.19
12.	129.41	66.05	34.02	17.97	9.66	4.92
Average Output IMF Frequencies						
Mag	114.0	62.9	32.8	16.9	9.3	5.1
Wrt Inp	2.3%	5.2%	10.4%	0.8%	2.9%	3.2%
Output IMF Frequency STDs						
Mag	11.2	7.3	4.0	2.2	1.3	0.7
Wrt Avg	9.8%	11.7%	12.1%	13.2%	13.9%	14.4%

(MIMO) signal analysis where in reality each signal is represented by its EMD. Recall from the transformed input-output IMFs, $\{Z_x(t), Z_y(t)\}$, the corresponding set of empirigrams $\{H_x(\eta_x, t), H_y(\eta_y, t)\}$ were defined in Eq. 7 for each set of $\eta = \eta_x = \eta_y$ common IMFs from the input and output EMDs. For the analysis described here it is not necessary to use an identical number, η, of IMFs from input and output (but this restriction is maintained here for simplicity). Then the Hilbert cross-empirigram $H_{xy}(\eta, t) = H_x^*(\eta, t)H_y(\eta, t)$ correlates the respective analytic IMFs from the two Hilbert empirigrams. Now define

$$\sigma_{xy}(\eta) \quad = \quad \sigma[H_x^*(\eta, t)H_y^T(\eta, t)]$$

where σ_{xy} represents the η-vector of singular values from the singular value decomposition (SVD) of the product of the empirigrams. The singular values represent relative contributions from the principal cross-correlation analytic IMFs as a result of correlation of all input analytic IMFs to all output analytic IMFs over the entire time span. Therefore, higher σ-valued cross-analytic IMFs have more input-output significance in terms of operator norm from input to output. The maximum singular value, $\bar{\sigma}_{xy} = \max_\eta(\sigma_{xy}(\eta))$, of this input-output operator corresponds to the structured singular value with a full-complex uncertainty block structure. In this context, for $M^{\eta \mathbf{x} \eta} = H_x^*(\eta, t)H_y^T(\eta, t)$, a complex matrix operator of input analytic IMFs to output analytic IMFs, $\mu_\Delta(M)$ is a measure of the smallest uncertainty

$\Delta = \mathbf{C}^{\eta \mathbf{x} \eta}$ (note that if $\eta_x \neq \eta_y$, then $\Delta = \mathbf{C}^{\eta_y \mathbf{x} \eta_x}$ is not square) that causes instability if interpreted as a constant matrix feedback loop [32],

$$\mu_{\{\Delta = \mathbf{C}^{\eta \times \eta}\}}(M) \quad = \quad \bar{\sigma}[H_x^*(\eta, t) H_y^T(\eta, t)]$$

and for scalar uncertainty structure $\Delta = \{\delta I_\eta : \delta \in \mathbf{C}\}$ representing diagonal structure between input-output IMFs,

$$\mu_\Delta(M) \quad = \quad \rho[H_x^*(\eta, t) H_y^T(\eta, t)]$$

where ρ is the spectral radius (largest magnitude eigenvalue). This is distinctly different from the full-block structure in that uncertainty is only between corresponding input-output analytic IMFs (similar dominant frequencies), while ignoring uncertainty across different analytic IMFs (of different dominant frequencies) between input and output. This is obviously less conservative but generally less realistic as well, especially for nonlinear effects which cross frequencies. Larger μ_Δ values in either case represent effects of uncertainty between input and output such that higher correlated IMFs relate to more sensitivity to uncertainty at those dominant IMF frequencies.

For MIMO signal analysis, it would be most appropriate to combine complex blocks (either full or scalar) for each input-output into a multi-block structure, where each complex sub-block corresponds to an input-output analytic IMF complex uncertainty structure. In computation this is often expanded in a block-diagonal context with repeated scalar and full blocks [32], where in this case each of these blocks would correspond to a single input-output analytic IMF uncertainty structure.

TABLE 4. Normalized μ_Δ results from F/A-18 AAW left and right wing acceleration data.

Full	3. L-wt-fwd	7. L-wof-fwd	11. L-wif-fwd	9. R-wif-fwd	5. R-wof-fwd	1. R-wt-fwd
Block	4. L-wt-aft	8. L-wof-aft	12. L-wif-aft	10. R-wif-aft	6. R-wof-aft	2. R-wt-aft
Fwd-wing	0.29	0.29	0.73	1.00	0.62	0.80
Aft-wing	0.65	0.67	0.56	0.19	0.31	0.72
Scalar	3. L-wt-fwd	7. L-wof-fwd	11. L-wif-fwd	9. R-wif-fwd	5. R-wof-fwd	1. R-wt-fwd
Block	4. L-wt-aft	8. L-wof-aft	12. L-wif-aft	10. R-wif-aft	6. R-wof-aft	2. R-wt-aft
Fwd-wing	0.11	0.13	0.19	1.00	0.47	0.49
Aft-wing	0.23	0.42	0.27	0.11	0.06	0.34

To compare input-to-output correlations from aileron position to wing accels for different uncertainty structures, structured singular values were computed for the full-block and scalar-block structures between the collective aileron position and F/A-18 AAW wing accelerometer responses, then normalized with respect to the largest of the group. Table 4 lists the results for both types of uncertainty structures. Accelerometer#9 has the highest correlation (=1.00) with the input aileron position in either case. Interestingly, comparison between left and right wing accels is poor even from a symmetric aileron input, thereby indicating asymmetry and/or nonlinearity. Lower μ_Δ values indicate a degree of robustness to uncertainty in a feedback stability sense, and as pointed out previously, scalar uncertainty is prevalently less conservative (as evidenced from absolute values before normalizing since

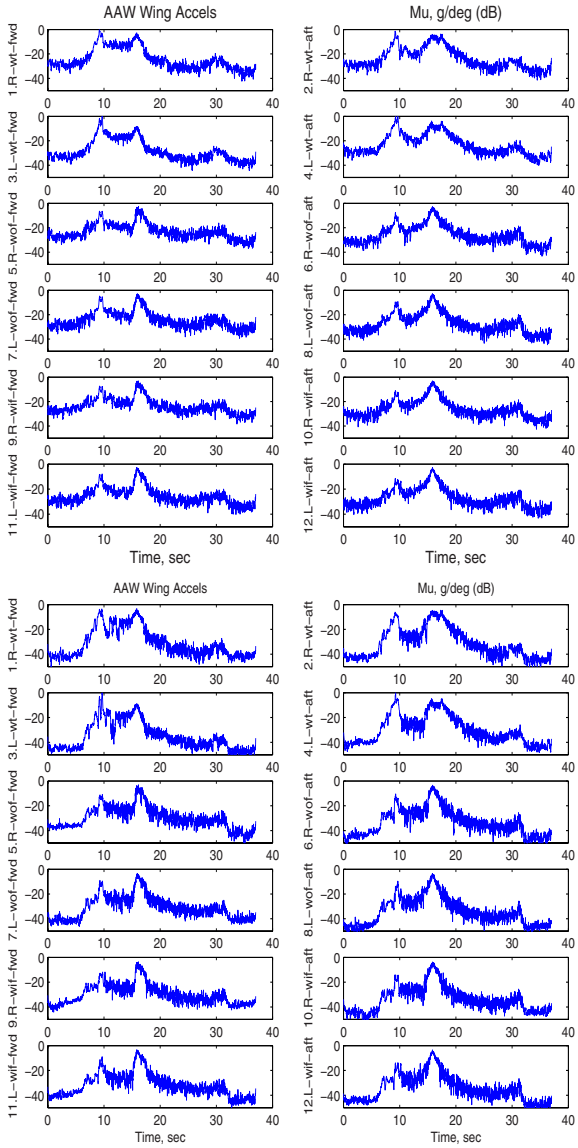

FIGURE 13. Time-varying $\mu_\Delta(M,t)$ (dB) for full-block (top plots) and scalar-block (bottom plots) uncertainty.

the normalization factor is almost equal between the uncertainty structures). If an a priori bound on gain levels (normalization) between input-output is defined such that $\| \Delta \|_\infty < 1$ is considered acceptable in some sense (stability, health diagnostic, safety margin), then μ_Δ has an absolute (scaled) interpretation such that $\mu_\Delta > 1$ indicates that an acceptable threshold has been exceeded, and this can be used as a health, stability, or safety monitor.

A time-dependent interpretation is also available by calculating $\mu_\Delta(M,t)$ at each time point.

$$\begin{aligned} \mu_\Delta(M,t) &= \rho[H_x^*(\eta,t_i)H_y^T(\eta,t_i)]\forall t_i \quad \text{(scalar-block)} \\ &= \bar{\sigma}[H_x^*(\eta,t_i)H_y^T(\eta,t_i)]\forall t_i \quad \text{(full-block)} \end{aligned}$$

In Fig. 13 are plots of $\mu_\Delta(M,t)$ for each input-to-output (aileron-to-accel) normalized analytic IMFs where it is evident that values close to one (0 dB) are near a unity operator norm limit. Again, this depends on the bound $\| \Delta \|_\infty < 1$ indicating an acceptable threshold, in this case maximum input-to-output gain normalized to one. Values close to one (0 dB) approach the acceptable limit. This uncertainty structure application has implications for model validation in the time domain [19].

7.2. Aerostructures Test Wing (ATW)

Another example which includes an actual instability is taken from Aerostructures Test Wing [12] (ATW) flight test data. The input is a sine sweep PZT voltage and outputs are three wingtip accelerometer responses. The EMDs of the PZT input and center wingtip accelerometer output near the flutter condition are displayed in Fig. 14. Flutter response EMDs of the center wingtip accelerometer and corresponding $\mu_\Delta(M,t)$ plots between input analytic IMFs and all three wingtip accelerometer output analytic IMFs are shown in Fig. 15. The input PZT was not activated during the flutter occurrence, so the input is essentially a small arbitrary oscillation about zero. It is evident that at the point of instability past 7.5 s, $\mu_\Delta(M,t)$ appropriately approaches a value of one (0 dB) since the input and output are normalized to unity, and assuming this limit corresponds to the limit of stability, and output response approaches its upper limit at flutter.

8. Parameter Estimation Using HHT

The Hilbert transform can used for estimating modal parameters such as natural frequencies and damping ratios. For a single-mode system, the damped natural frequency can be easily determined from the Hilbert transform of the impulse response function. In this case, the damped natural frequency ω_d is given by the slope of the instantaneous phase angle plotted as a function of time, and damping ratio ζ is estimated in a straightforward manner [24].

For a linear single-mode system characterized by a pair of complex conjugate eigenvalues, the eigenvalues are directly related to the modal parameters as $-\zeta\omega_n + \omega_d$ should be imaginary, where ω_n denotes the natural frequency. The impulse response function is

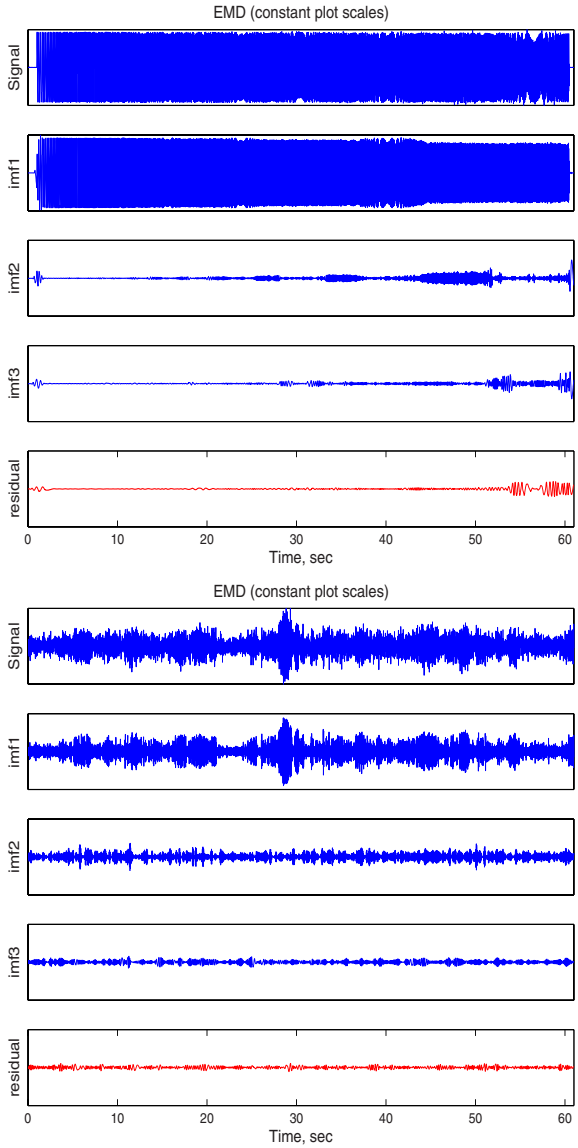

FIGURE 14. Empirical mode decomposition of the ATW input PZT (left plot) and center wingtip accelerometer (right plot) at Mach 0.82, 10,000 feet (3,048 m) altitude, just before flutter (original signal at top and residual at bottom in each set).

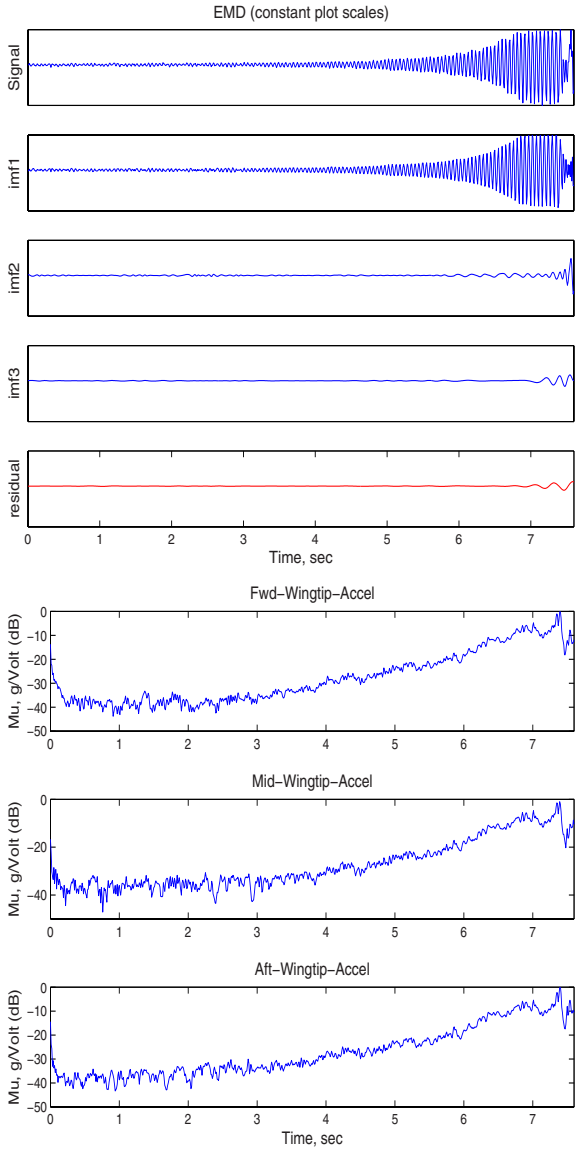

FIGURE 15. Empirical mode decomposition of the center wingtip accelerometer (left plot) near Mach 0.83, 10,000 feet (3,048 m) altitude, at flutter (original signal at top and residual at bottom), and time-varying $\mu_\Delta(M,t)$ for full-block uncertainty between input and output IMFs of the three wingtip accelerometer outputs (right plots).

$$h(t) = \frac{1}{m\omega_d} \mathrm{e}^{-\zeta\omega_n t} \sin(\omega_d t)$$

where m denotes the mass. Provided the damping ratio is small, the following relationships hold [24]

$$\begin{aligned} \mathrm{H}\left\{\mathrm{e}^{-\zeta\omega_n t}\sin(\omega_d t)\right\} &= \mathrm{e}^{-\zeta\omega_n t}\cos(\omega_d t) \\ \mathrm{H}\left\{\mathrm{e}^{-\zeta\omega_n t}\cos(\omega_d t)\right\} &= -\mathrm{e}^{-\zeta\omega_n t}\sin(\omega_d t) \end{aligned}$$

where again H denotes the Hilbert transform. Therefore, the Hilbert transform of the impulse response function can be written as

$$\mathrm{H}\left\{h(t)\right\} := \tilde{h}(t) = \frac{1}{m\omega_d}\mathrm{e}^{-\zeta\omega_n t}\cos(\omega_d t).$$

The impulse response function can then be viewed as the real part of the following analytic signal [24]

$$\begin{aligned} z(t) &= h(t) + \mathrm{i}\tilde{h}(t) \\ &= \frac{1}{m\omega_d}\mathrm{e}^{-\zeta\omega_n t}\sin(\omega_d t) + \mathrm{i}\frac{1}{m\omega_d}\mathrm{e}^{-\zeta\omega_n t}\cos(\omega_d t). \end{aligned}$$

The amplitude of this signal is given by

$$a(t) = \sqrt{h(t)^2 + \tilde{h}(t)^2} = \frac{1}{m\omega_d}\mathrm{e}^{-\zeta\omega_n t}$$

which represents the envelope of the impulse response. Taking the natural logarithm of the amplitude yields

$$\ln\left(a(t)\right) = -\zeta\omega_n t - \ln\left(m\omega_d\right).$$

The slope σ of the natural log of $a(t)$ plotted against time gives $\sigma = -\zeta\omega_n$. Given that ω_d can be obtained from the slope of the phase plotted against time, and using the definition of the damped natural frequency, there results ω_n,

$$\begin{aligned} \omega_d &= \sqrt{1-\zeta^2}\omega_n \\ \omega_d^2 &= (1-\zeta^2)\omega_n^2 = \omega_n^2 - \sigma^2 \\ \omega_n &= \sqrt{\omega_d^2 + \sigma^2} \end{aligned}$$

and the damping ratio can be calculated as $\zeta = \sigma/\omega_n$. This approach to damping estimation is in the same spirit as the logarithmic decrement technique commonly used in structural dynamics.

Most systems have multiple modes, however, which implies that the impulse response function will generally contain contributions from several modes. Therefore, because they are valid only for single-mode systems, logarithmic decrement or the approach based on the Hilbert transform cannot be used for estimating damping. As emphasized earlier, the Hilbert transform does not yield meaningful frequency information for multiple-mode systems since it attempts to identify a single instantaneous frequency at each time step. The Hilbert-Huang algorithm makes

it possible to apply the above technique for damping and frequency estimation because it can decompose a multi-component signal into a series of single-component signals through the EMD process. By performing EMD on the impulse response function of a multiple-mode system, the contributions of the different modes can be extracted and analyzed separately. This approach has been applied by Yang et al. [27, 28, 29] in a series of papers in which the Hilbert-Huang algorithm was used for parameter estimation of several multiple degree-of-freedom structures. A similar analysis is used here to estimate damping ratios and frequencies for aeroelastic systems.

As an example, consider a prototypical linear pitch-plunge aeroelastic system that has been studied extensively in the literature. The system consists of an airfoil with pitch and plunge degrees of freedom, and the input is the deflection of the trailing edge flap. In this study, the flow velocity U is varied until the linear flutter speed is reached at approximately $U = 11.8$ m/s.

Because this system has two outputs, pitch and plunge, there are two impulse response functions that can be measured. Each of these functions generally contains contributions from both modes of the system. As an example, Fig. 16 depicts the pitch and plunge impulse response functions for a flow velocity of $U = 8$ m/s. These were obtained by simulating the pitch and plunge responses to an impulse applied to the trailing edge flap. The plunge impulse responses are used over a range of flow velocities to measure frequencies and damping ratios using the Hilbert-Huang algorithm. These impulse response functions are obtained by way of simulation in this paper, but in practice they can be obtained by way of system identification techniques such as taking the inverse Fourier transform of the measured frequency response function or identifying first-order Volterra kernels from measured input-output data.

The plunge impulse response functions were used to estimate the damped natural frequencies, damping ratios, and natural frequencies over a range of flow velocities. The results are summarized in Table 5, which lists the actual and estimated damped natural frequencies and damping ratios for flow velocities ranging from $U = 1$ m/s to $U = 11.9$ m/s (just past the linear flutter speed). The actual frequencies and damping ratios were calculated by solving for the eigenvalues of the system. The estimation of the modal parameters for $U = 8$ m/s is demonstrated in Figs. 17 and 18. Figure 17 illustrates the EMD of the plunge impulse response function at $U = 8$ m/s. In this case, both modes of vibration are evident in the first intrinsic mode function. This is because the first, higher-frequency mode damps out quickly and the second mode persists for a longer period of time. Therefore, the frequencies and damping ratios of both modes were estimated by using different portions of the first IMF, as depicted in Fig. 18. The figure shows plots of the instantaneous phase and the natural log of the amplitude of the first IMF. Recall that the slopes of these lines are used to estimate the modal parameters. The initial portion of the amplitude plot is not a straight line because of initial transients in the IMF and possible boundary effects. Therefore, the portions of the plot that most resembled straight lines were used for the slope calculations

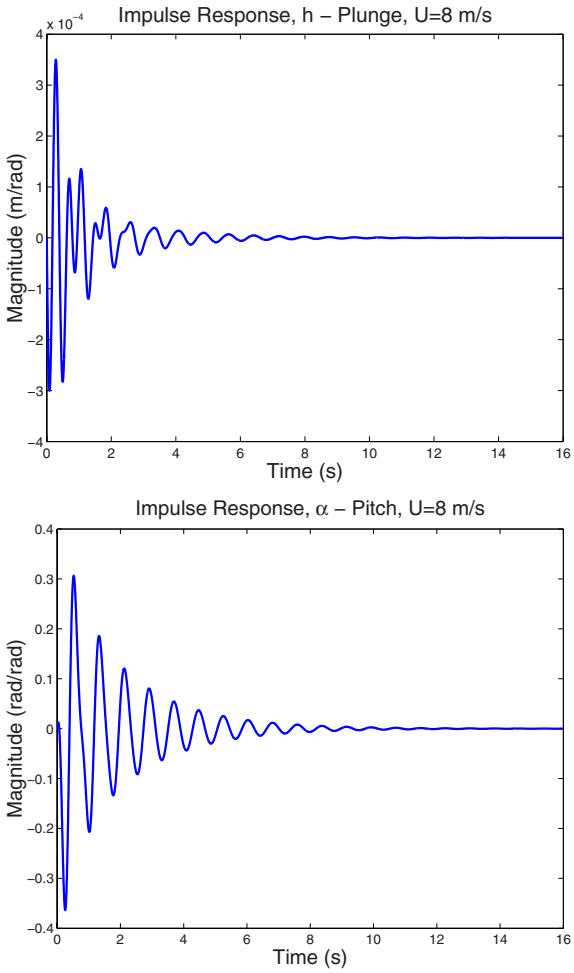

FIGURE 16. Plunge and pitch impulse response functions for $U = 8$ m/s

for each mode, as labeled in the figure. Because this analysis technique is only approximate depending on the magnitude of the damping ratio (increased error for larger damping), the amplitude curve tends to oscillate about an average slope. Therefore, the slope in each case was obtained using a linear least-squares curve-fit of the data [27, 28, 29].

The results presented in Table 5 show that the estimates of the damped natural frequencies and damping ratios are fairly accurate in most cases. As the linear flutter speed (approximately $U = 11.8$ m/s) is approached, the plunge mode is not discernible in the IMFs. In this regime, the system behaves essentially as a single-mode system and it was not possible to estimate the damping and frequency

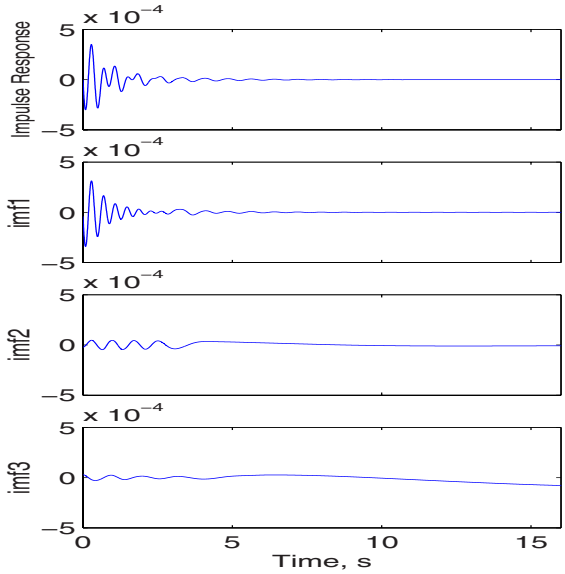

FIGURE 17. Empirical mode decomposition of the plunge impulse response function (m/rad) at $U = 8$ m/s.

of the plunge mode. The actual and estimated damping ratios for each mode are plotted as a function of flow velocity in Fig. 19.

A few observations are now presented. The first IMF was used to estimate the parameters of both modes. The analysis would be somewhat cleaner if each mode appeared in a different IMF. This is indeed the case in the work of Yang et al. [27, 28, 29] in which the frequency range of each IMF is controlled in the EMD process. Therefore, controlling the frequency range of each IMF is an option that could enhance the parameter estimation procedure. Also, even though parameters for both modes were estimated using a single IMF, the same results would generally not be obtained from an analysis of the impulse response function without the benefit of EMD. Although the IMF used in the analysis contains contributions from two modes, only one mode is present in the signal at any point in time. In contrast, the impulse response function may contain contributions from both modes over the same time period. This example demonstrates some potential benefits of the HHT for extracting modal parameters from aeroelastic data.

9. Conclusions

Application of the Hilbert-Huang algorithm for system signal decompositions, studying the effect of enhancements such as local/on-line behavior, understanding

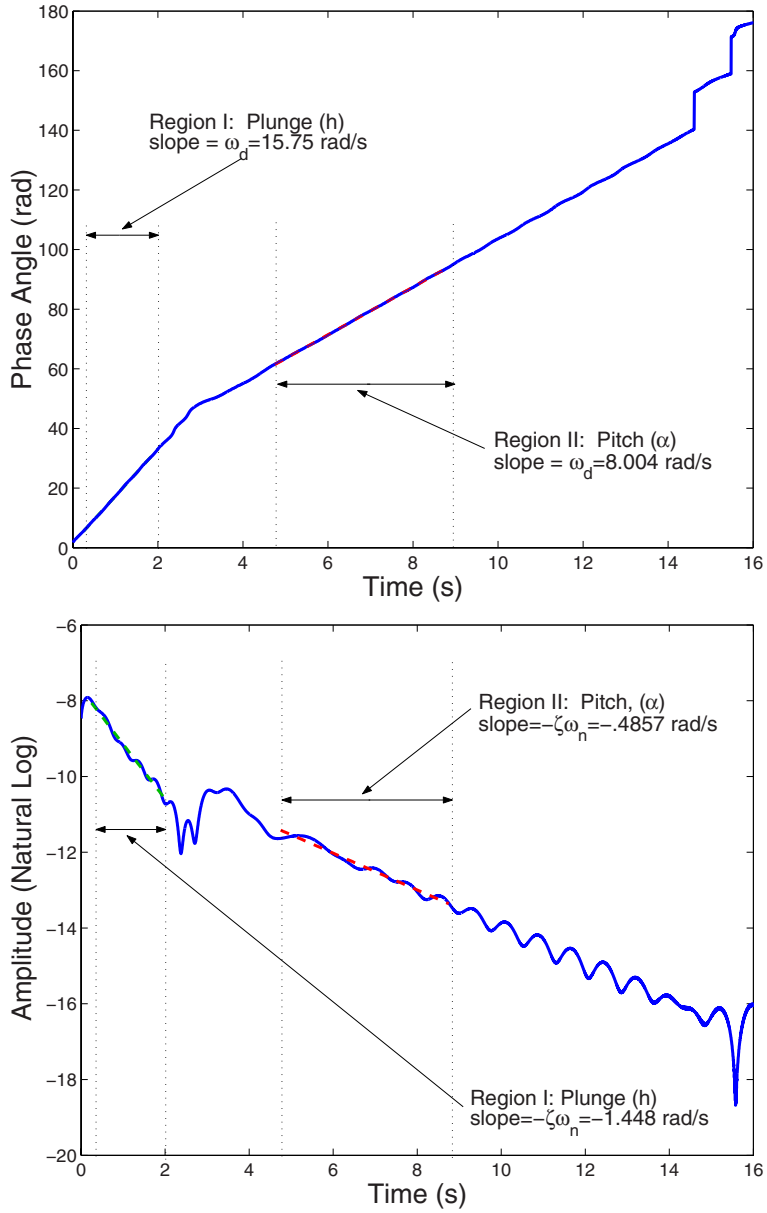

FIGURE 18. Instantaneous phase and amplitude plots used to estimate modal parameters from the first IMF of the plunge impulse response function at $U = 8$ m/s.

TABLE 5. Actual and estimated (Est) damped frequencies (Hz) and damping ratios.

Velocity (m/s)		ω_d	Est ω_d	ζ	Est ζ
1	Plunge	2.765	2.735	.0854	.0874
	Pitch	1.027	1.022	.0441	.0444
3	Plunge	2.739	2.709	.0855	.0881
	Pitch	1.057	1.050	.0503	.0490
5	Plunge	2.684	2.655	.0862	.0895
	Pitch	1.119	1.096	.0553	.0591
7	Plunge	2.595	2.567	.0877	.0926
	Pitch	1.217	1.214	.0590	.0579
9	Plunge	2.454	2.452	.0906	.0957
	Pitch	1.366	1.362	.0607	.0637
11	Plunge	2.198	–	.1012	–
	Pitch	1.628	1.619	.0544	.0544
11.5	Plunge	2.073	–	.1174	–
	Pitch	1.755	1.744	.0392	.0394
11.6	Plunge	2.040	–	.1262	–
	Pitch	1.788	1.775	.0305	.0305
11.7	Plunge	2.009	–	.1392	–
	Pitch	1.819	1.806	.0176	.0177
11.8	Plunge	1.984	–	.1552	–
	Pitch	1.844	1.831	.0016	.0017
11.9	Plunge	1.967	–	.1713	–
	Pitch	–	–	-.0145	-.0146

filtering properties, and especially for investigating correlations between input-output and between system sensors in terms of instantaneous properties, is revealed. System input-output signal analysis is introduced to characterize the time-varying amplitude and frequency components of multiple data channels, including input-to-output and distributed sensors, in terms of the intrinsic mode functions (IMFs) of the Hilbert-Huang transform (HHT). Significant departures from Fourier and other time-frequency or time-scale wavelet approaches are exposed. In these respects, this paper attempts to show how the HHT behaves in a sometimes non-intuitive and subtle manner in the analysis of F/A-18 Active Aeroelastic Wing (AAW) aircraft aeroelastic flight test data. Online stability analyses and modal identification are also presented. Examples are given using aeroelastic test data from the F/A-18 Active Aeroelastic Wing aircraft, an Aerostructures Test Wing, and pitch-plunge simulation.

An objective signal-adaptive basis function derivation, the Hilbert-Huang algorithm yields intrinsic mode functions giving instantaneous frequencies as functions of time that permit identification of embedded structures. There is a multiresolution quality in the empirical mode decomposition process which even deals

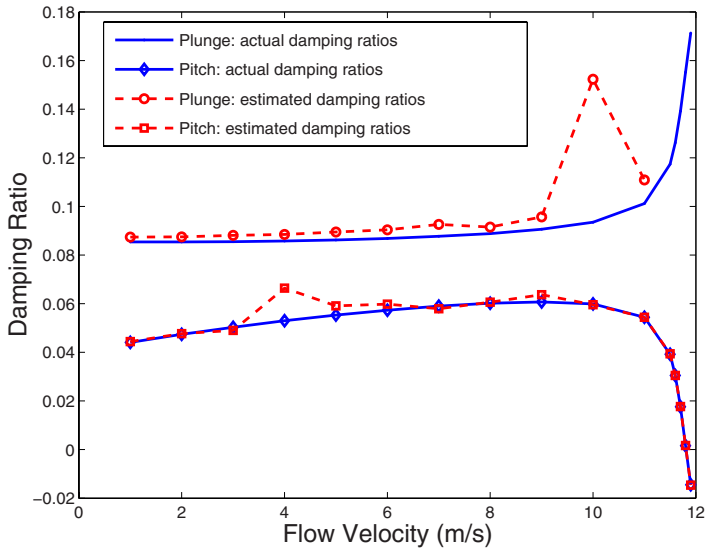

FIGURE 19. Actual and estimated damping ratios.

with intermittency by allowing multiple time-scales within an intrinsic mode function, but not allowing a similar time-scale simultaneously with other IMFs. System identification in the IMF sub-component environment is a practical endeavor in the domain of multiresolution system identification. It should be noted that the idea of exploiting local properties for signal analysis applies to spatial data as well as temporal data with frequency and scale (translation and duration) variations. From the idea of empiquency to describe oscillations in images based on extrema points there are potential applications for general time-space-frequency-scale signal processing.

Modern intelligent control and integrated aerostructures require control feedback signal processing cognizant of system stability and health. Time-varying linear or nonlinear modal characteristics derived from flight data are all within the realm of HHT. Further research will investigate these issues and HHT connections between localized instantaneous dynamics, health diagnostics, and global system stability and performance for monitoring and prediction.

References

[1] Bendat, J. S. and A.G. Piersol, *Random Data Analysis and Measurement Procedures*, John Wiley & Sons, Inc., Chap. 12-13, 2000.

[2] Buresti, G., G. Lombardi, and J. Bellazzini, "On the analysis of fluctuating velocity signals through methods based on the wavelet and Hilbert transforms," *Chaos, Solitons and Fractals*, **20**(2004), pp. 149-158.

[3] Chen, Q., N. Huang, S. Riemenschneider, and Y. Xu, "A B-spline approach for empirical mode decompositions," *Advances in Computational Mathematics*, 2004.

[4] Cohen, L., "Time-Frequency Distributions - A Review," *Proc. IEEE*, **77**(7), 1989, pp. 941-981.

[5] Flandrin, P., *Time-Frequency/Time-Scale Analysis*, Academic Press, 1999.

[6] Flandrin, P., G. Rilling, and P. Gonçalvès, "Empirical Mode Decomposition as a Filter Bank," *IEEE Signal Processing Letters*, **11**(2), Feb 2004, pp. 112-114.

[7] Flandrin, P. and P. Gonçalvès, "Empirical Mode Decompositions as Data-Driven Wavelet-Like Expansions," accepted for publication in *International Journal of Wavelets, Multiresolution and Information Processing*, 2005.

[8] Huang, N. E., Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N-C Yen, C.C. Tung, and H.H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Royal Society London A*, **454**(1998), pp. 903-995.

[9] Huang, N. E., Z. Shen, and S.R. Long, "A New View of Nonlinear Water Waves: The Hilbert Spectrum," *Annual Review of Fluid Mechanics*, **31**(1999), pp. 417-457.

[10] Huang, N. E., M-L.C. Liu, S.R. Long, S.S.P. Shen, W. Qu, P. Gloersen, and K.L. Fan, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," *Proc. Royal Society London A*, **459**(2003), pp. 2317-2345.

[11] Jha, R., F. Yan, and G. Ahmadi, "Energy-Frequency-Time Analysis of Structural Vibrations using Hibert-Huang Transform," AIAA-2004-1975.

[12] Lind, R., D.F. Voracek, R. Truax, T. Doyle, S. Potter, and M. Brenner, "A flight test to demonstrate flutter and evaluate the flutterometer," *The Aeronautical Journal*, Oct 2003.

[13] Linderhed, A., "2D empirical mode decompositions in the spirit of image compression," *Wavelet and Independent Component Analysis Apps IX, SPIE Proc.*, **4738**(2002), pp. 1-8.

[14] Linderhed, A., "Image compression based on empirical mode decomposition," *Proc. of SSAB 04 Symp. on Image Analysis*, Uppsala, Sweden, Mar 11-12, 2004, pp. 110-113.

[15] Liu, B., S. Riemenschneider, and Y. Xu, "Gearbox fault diagnosis using empirical mode decomposition and Hilbert spectrum," submitted for publication in *MSSP*, 2004.

[16] Liu, Z. and S. Peng, "Boundary Processing of Bidimensional EMD Using Texture Synthesis," *IEEE Signal Processing Letters*, **12**(1), Jan 2005, pp. 33-36.

[17] Loughlin, P. J. and L. Cohen, "The Uncertainty Principle: Global, Local, or Both?," *IEEE Transactions on Signal Processing*, **52**(5), May 2004, pp. 1218-1227.

[18] Pendleton, E., D. Bessette, P. Field, G. Miller, and K. Griffin, "Active Aeroelastic Wing Flight Research Program: Technical Program and Model Analytical Development," *AIAA Journal of Aircraft*, **37**(4), Jul-Aug 2000, pp. 554-561.

[19] Poolla, K., P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal, "A Time-Domain Approach to Model Validation," *IEEE Transactions on Automatic Control*, **39**(5), May 1994, pp. 951-959.

[20] Qian, Tao, Q. Chen, and L. Li, "Analytic Unit Quadrature Signals with Nonlinear Phase", Physica D: Nonlinear Phenomena, Vol. 203, Issues 1-2, pp. 80-87 (2005).

[21] Rilling, G., P. Flandrin, and P. Gonçalvès, "On Empirical Mode Decomposition and its algorithms," *IEEE-EURASIP Workshop on Nonlinear Signal/Image Processing*, Grado, Italy, 2003.

[22] Sharpley, R. C. and V. Vatchev, "Analysis of the Intrinsic Mode Functions," submitted for publication in *Constructive Approximations*, 2005.

[23] Vatchev, V., *Intrinsic Mode Functions and the Hilbert Transform*, Ph.D thesis, Dept of Mathematics, U. of South Carolina, 2004.

[24] Worden, K. and G.R. Tomlinson, *Nonlinearity in Structural Dynamics: Detection, Identification, and Modeling*, Institute of Physics Publishing Ltd, Bristol and Philadelphia, 2001.

[25] Wu, Z., N.E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. Royal Society London A*, **460** (2004), pp. 1597-1611.

[26] Yang, B. and C.S. Suh, "Interpretation of crack-induced rotor non-linear response using instantaneous frequency," *Mechanical Systems and Signal Processing*, **18**(2004), pp. 491-513.

[27] Yang, J. N., Y. Lei, S. Pan, and N. Huang, "System identification of linear structures based on Hilbert-Huang spectral analysis. Part 1: Normal modes," *Earthquake Engineering and Structural Dynamics*, **32**(2003), pp. 1443-1467.

[28] Yang, J. N., Y. Lei, S. Pan, and N. Huang, "System identification of linear structures based on Hilbert-Huang spectral analysis. Part 2: Complex modes," *Earthquake Engineering and Structural Dynamics*, **32**(2003), pp. 1533-1554.

[29] Yang, J. N., Y. Lei, S. Pan, and N. Huang, "Identification of Natural Frequencies and Dampings of In Situ Tall Buildings Using Ambient Wind Vibration Data ," *Journal of Engineering Mechanics*, **130**(1), May 2004, pp. 1-8.

[30] Yang, J. N., Y. Lei, S. Pan, and N. Huang, "Hilbert-Huang Based Approach for Structural Damage Detection," *Journal of Engineering Mechanics*, **130**(1), 2004, pp. 85-95.

[31] Yang, Z., D. Qi and L. Yang, "Signal Period Analysis Based on Hilbert-Huang Transform and Its Application to Texture Analysis," *Third Intl Conf Image/Graphics, ICIG'04*, pp. 430-433.

[32] Zhou, K., with J. Doyle, *Essentials of Robust Control*, Prentice Hall, New Jersey, 1998.

Martin J. Brenner
Aerospace Engineer, Aerostructures Branch
NASA Dryden Flight Research Center
Edwards, CA 93523-0273, USA
e-mail: `Martin.J.Brenner@nasa.gov`

Sunil L. Kukreja
Aerospace Engineer, Aerostructures Branch
NASA Dryden Flight Research Center
Edwards, CA 93523-0273, USA
e-mail: `Sunil.Kukreja@nasa.gov`

Richard J. Prazenica
Visiting Assistant Professor
UF Research in Engineering Educational Facility (REEF)
University of Florida
Shalimar, FL 32579, USA
e-mail: prazenic@ufl.edu

Wavelet Analysis and Applications
Tao Qian, Mang I. Vai and Xu Yuesheng, Eds.
Applied and Numerical Harmonic Analysis, 363–376

An Adaptive Data Analysis Method for Nonlinear and Nonstationary Time Series: The Empirical Mode Decomposition and Hilbert Spectral Analysis

Norden E. Huang

Abstract. An adaptive data analysis method, the Empirical Mode Decomposition and Hilbert Spectral Analysis, is introduced and reviewed briefly. The salient properties of the method is emphasized in this review; namely, physical meaningful adaptive basis, instantaneous frequency, and using intra-wave frequency modulation to represent nonlinear waveform distortion. This method can perform and enhance most of the traditional data analysis task such as filtering, regression, and spectral analysis adaptively. Also presented are the mathematical problems associated with the new method. It is hope that this presentation will entice the interest of the mathematical community to examine this empirically based method and inject mathematical rigor into the new approach.

Mathematics Subject Classification (2000). Primary 99Z99; Secondary 00A00.

1. Introduction

Data analysis is necessary for science and engineering, for data is the only link we have with the reality. Consequently, data analysis serves two purposes: First, it provides validation of our theories or models. Second, it provides the guide of the underlying mechanisms as a base for discovery, creation or improvements of the theories and models. Either way, the data contains information we are seeking; the goal of data analysis is to find the information in the data. The common tasks involved in data analysis are finding the distributions, filtering, time-frequency, regressions and error analyses. Even on the limited scope of filtering to cleanse the data of unwanted noise, we still face with the daunting task of determining which part of the data is noise and which part is valid information. In the filtering process we should only eliminate the noise, but not degrade the information. As we do not

have a complete knowledge base of the underlying mechanisms for most of the physical problems we face today, we should inject as little subjective specifications as possible in the process of data analysis, so that we do not prejudice the results. A truly objective data analysis method should be adaptive to the data and let the data set speaks for itself.

Traditional time-frequency analysis methods, however, all follows the well established mathematical rules: the methods all start with a definition of basis, and convolve the signal with the basis to get amplitude and frequency either for distributions or for filtering. Such an approach has the great advantage of having a solid mathematical foundation. Once the algorithm is established, data analysis can go forward mechanically. Unfortunately, within the comfortable fold of solid mathematic foundation, the methods can not be adaptive at all. Furthermore, this well trodden path also restricts the methods developed under this paradigm to linear and stationary assumptions.

As data can come from all sources ranging from relatively well established physical sciences, to complicated biologic processes and social-economic phenomena., most of the driving mechanisms are so complicatedly intertwined and interacting that the data we obtained are also highly variable, not only from one case to another but also from time to time even limited to one single case. In other words, we have to face data from nonlinear and nonstationary processes. This requirement is known for a long time, but remedy is slow to come. To accommodate for data from nonstationary processes, we have met more success. Methods (see for example, Flandrin, 1999) such as spectrogram, Wigner-Ville distribution, Wavelet analysis are all examples. To accommodate for data from nonlinear processes, however, our progress has been very slow. The available methods (see, for example, Tong, 1990, Krantz and Schreiber, 1997 and Diks, 1998) are limited to handle data from deterministic low dimensional chaotic systems.

Even for data from nonstationary processes, the available methods are also limited to linear systems, for the methods were mostly based on the well established *a priori* basis approach, where all the analysis is based on convolution of the data with the established basis. This powerful mathematical tool, unfortunately, has drained all the physics out of the analyzed results, for any *a priori* basis could not possibly fit all the variety of data from drastically different underlying driving mechanisms. Any misfit will automatically be assigned to the various orders of harmonics with respect to the selected basis. Though results so obtained satisfy the mathematical requirements, they lack physical meaning. Furthermore, the convolution processes involve integration, which make the results suffering the limitation imposed by the uncertainty principle, and preventing us from examine the details of the data and their underlying mechanisms.

Let us take a simple example to examine the characteristics of data from a nonlinear system. Consider the Duffing equation without damping given as

$$\frac{d^2x}{dt^2} \; + \; x \; + \; \varepsilon x^3 \; = \; \gamma \cos \omega t \; , \tag{1.1}$$

where ε is a parameter, not necessarily small; γ is the magnitude of the driving force. We can easily rewrite this equation slightly as follows:

$$\frac{d^2x}{dt^2} \; + \; x \; (1 \; + \; \varepsilon x^2 \,) \; = \; \gamma \cos \omega t \; . \tag{1.2}$$

If we treat the quantity in the parenthesis as a single number designated as L:

$$L \; = \; 1 \; + \; \varepsilon \, x^2, \tag{1.3}$$

then the quantity L can be treated as the pendulum length or the spring constant. Either way, L changes with position; therefore, the frequency of the system should also change with position even within one oscillation period. Such intra-wave frequency modulation is the special characteristics of a nonlinear system; and it requires a detailed frequency representation that is unattainable form *a priori* basis approach. For example, following the classic perturbation analysis by imposing a linear structure on a nonlinear system, one would find the solution consisted of endless harmonics. The effect of the harmonics is to find enough sinusoidal components to fit the deformed final waveform, commonly known as harmonic distortions. It is well know that each term in this perturbation solution does not have physical meaning, only the totality of all the terms represents the physics. But using any *a priori* basis analysis, one would inevitably obtain a collection of the harmonics of one form or the other depending on the basis function selected; thus rendered the interpretation of spectral analysis problematical. The harmonics representation here is a poor substitute of the detailed instantaneous frequency description of the intra-wave frequency modulation. But such a detailed description will call for a drastic new approach. In fact to describe intra-wave frequency modulation, we cannot use *a priori* basis approach. An easy alternative is to use the Hilbert Transform, which is defined as

$$y(t) \; = \; \frac{1}{\pi} \; P \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} \; , \tag{1.4}$$

in which *x(t)* is the given function of *Lp* class, *y(t)* is the Hilbert transform, which is the complex conjugate of x(t), and *P* indicates the principal value of the singular integral. As y(t) is the complex conjugate, we have

$$z(t) \; = x \, (t) \; + \; j \; y(t) \; = \; a \, (t) \, e^{j \, \theta(t)} \; , \tag{1.5}$$

where

$$a(t) \; = \; \left(x^2 \; + \; y^2\right)^{1/2} \quad ; \quad \theta(t) \; = \; \tan^{-1} \frac{y}{x} \; . \tag{1.6}$$

Here a is the instantaneous amplitude, and θ is the phase function; thus the instantaneous frequency, with the stationary phase approximation, is simply

$$\omega \;=\; \frac{d\theta}{dt}\;. \tag{1.7}$$

This definition also coincides with the classical wave theory. This definition of instantaneous frequency is indeed local, for it is defined through differentiation rather than integration. Therefore, the resulting instantaneous frequency should be able to describe the intra-wave frequency modulation. This approach has been recommended by Hahn (1996) for applications signal processing. Unfortunately, this straightforward and simple-minded approach does not work at all. Although the Hilbert transform is valid under a very general condition, for the instantaneous frequency derived from the above approach to make physical sense, the function has to be 'mono-component' as discussed by Cohen (1995) and Huang et al. (1998, 1999). This has been illustrated by Huang et al (1998) with a simple function as

$$x(t) \;=\; a \;+\; \cos\,\alpha t, \tag{1.8}$$

with a as an arbitrary constant. Its Hilbert transform is simply

$$y(t) \;=\; \sin\,\alpha t\,; \tag{1.9}$$

therefore, the instantaneous frequency according to Equation (7) is

$$\omega \;=\; \frac{\alpha\;(1\;+\;a\,\sin\,\alpha t)}{1+2a\,\cos\,\alpha t\;+\;a^2}\;\;. \tag{1.10}$$

Equation (10) can give any value for the instantaneous frequency, depending on the value of a. In order to recover the frequency of the input sinusoidal signal, the constant has to be zero. This simple example illustrates a crucial condition for the Hilbert Transform approach to work here: The function will have to be zero mean locally. This seemingly trivial condition has created great misunderstanding, which has prompted Cohen (1995) to list a number of 'paradoxes' concerning instantaneous frequency. Some of the paradoxes concerning negative frequency are direct consequence of this condition.

Another obvious consequence of this condition is the difficult experience by all previous attempts to use the Hilbert transform: how to reduce or decompose an arbitrary function to a 'mono-component' one with local zero mean? And more fundamentally, if the function is nonstationary, how can one find the local mean? These difficulties have forced the past applications of Hilbert transform to extract a narrow band component with a band-pass filter on the original data (Melville, 1983). As the band-pass filter is a linear operator, any signal passing through it will lost all its 'harmonics', and suffer deformation of the fundamental wave shape. Such approach certainly satisfies the condition demanded by the instantaneous frequency computation through Hilbert transform; it has unwittingly drained some interesting information form the data, the nonlinear characteristics associated with the signal.

With all these difficulties, the real applications of Hilbert transform will have to wait for the development of the Empirical Mode Decomposition (EMD) (Huang et al. 1998, 1999, 2003). Together with the Hilbert Spectral Analysis (HSA), the combination established a new adaptive time-frequency analysis method.

2. The Empirical Mode Decomposition and Hilbert Spectral Analysis

The details of both Empirical Mode Decomposition (EMD) and the Hilbert Spectral Analysis (HSA) are given in Huang et al. (1996, 1998 and 1999). The following summary is based on a simplified version given in Huang (2005). The EMD method is necessary to reduce any data from nonstationary and nonlinear processes into simple oscillatory function that will yield meaningful instantaneous frequency through the Hilbert transform. Contrary to almost all the previous decomposing methods, EMD is empirical, intuitive, direct, and adaptive, with the *a posteriori* defined basis based on and derived from the data. The decomposition is designed to seek the different simple intrinsic modes of oscillations in any data based on the principle of scale separation. The data, depending on it complexity, may have many different coexisting modes of oscillation at the same time. Each of these oscillatory modes is represented by an Intrinsic Mode Function (IMF) satisfying the following conditions:

(a) in the whole data set, the number of extrema and the number of zero-crossings must either equal or differ at most by one, and

(b) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The IMF is a counter part to the simple harmonic function, but it is much more general: instead of constant amplitude and frequency, IMF can have both variable amplitude and frequency as functions of time. This definition is inspired by the simple example of constant plus sinusoidal function given above. The total number of the IMF components is limited to $ln_2 N$, where N is the total number of data points. It satisfies all the requirements for a meaningful instantaneous frequency through Hilbert transform.

Pursuant to the above definition for IMF, one can implement the needed decomposition of any function, known as sifting, as follows: Take the test data; identify all the local extrema; divide the extrema into two sets: the maxima and the minima. Then connect all the local maxima by a cubic spline line to form an upper envelope. Repeat the procedure for the local minima to form a lower envelope. The upper and lower envelopes should encompass all the data between them. Their mean is designated as m_1, and the difference between the data and m_1 is designated as, h_1, a proto-IMF:

$$X(t) - m_1 = h_1 . \tag{2.1}$$

Ideally, *h1* should satisfy the definition of an IMF by construction of *h1* described above, which should have made it symmetric and having all maxima positive and all minima negative. Yet, in changing the local zero from a rectangular to a curvilinear coordinate system some inflection points could become additional extrema. New extrema generated this way actually reveal the hidden modes missed in the initial treatment. The sifting process sometimes can recover signals representing low amplitude riding waves with repeated siftings.

The sifting process serves two purposes: to eliminate riding waves, and to make the wave profiles more symmetric. While the first condition is absolute necessary for Hilbert transform to give a meaningful instantaneous frequency, the second condition is also necessary in case the neighboring wave amplitudes having too large a disparity. As a result, the sifting process has to be repeated many times to reduce the extracted signal an IMF. In the subsequent sifting process, *h1* is treated as the data for the next round of sifting; therefore,

$$h_1 \; - \; m_{11} \; = \; h_{11} \; . \tag{2.2}$$

After repeated sifting, up to k times, h_{1k}:

$$h_{1\,(k-1)} \; - m_{1k} \; = \; h_{1k} \; . \tag{2.3}$$

If h_{1k} becomes an IMF, it is designated as c_1:

$$c_1 \; = \; h_{1k} \; , \tag{2.4}$$

the first IMF component from the data. Here we have a critical decision to make: when to stop. Too many rounds of sifting will reduce the IMF to FM page criterion; to few rounds of sifting will not have a valid IMF. In the past, two different criteria have been used: The first one was used in Huang et al. (1998), based on a Cauchy type of convergence test. Specifically, the test is to require the normalized squared difference between two successive sifting operations defined as

$$SD_k \; = \; \frac{\sum\limits_{t=0}^{T} |h_{k-1}(t) \; - \; h_k(t)|^2}{\sum\limits_{t=0}^{T} h_{k-1}^2(t)} \; , \tag{2.5}$$

to be small. If this squared difference, SD_k, is small than a predetermined value, the sifting process will be stopped. Though this criterion is rigorous, it is very difficult to implement, for this criterion does not depend on the definition of the IMFs. The squared difference might be small, but there is no guarantee that the function will have the same numbers of zero-crossings and extrema, a necessary for a meaningful instantaneous frequency through Hilbert transform. These shortcomings prompted Huang et al. (1999 and 2003) to propose an alternative based on the agreement

of the numbers of zero-crossings and extrema. Specifically, an S-number is pre-selected. The sifting process will stop only if S consecutive times the numbers of zero-crossings and extrema stay the same, and are equal or at most differ by one. This second choice is not without its difficulty: how to select the S number. Obviously, any selection is *ad hoc*, and a rigorous justification is needed. But in Huang et al. (2003), they found the optimal S number to be around 4 to 8 for a variety of data sets. Such a selection could be used as a guide. Even if one chooses to eschew such an arbitrary selection, and ensemble of S number could give an ensemble mean of the various selection as shown in Huang et al (2003), which is even more meaningful.

With either stoppage criterion, the, *c1*should contain the finest scale or the shortest period component of the signal. We can, then, remove *c1*from the rest of the data by

$$X(t) - c_1 = r_1 . \tag{2.6}$$

Since the residue, r_1, contains all longer period variations in the data, it is treated as the new data and subjected to the same sifting process as described above. This procedure can be repeated to all the subsequent r_j's, and the result is

$$\begin{aligned} r_1 - c_2 &= r_2 , \\ &\cdots \\ r_{n-1} - c_n &= r_n \end{aligned} \tag{2.7}$$

The sifting process should stop when the residue, r_n, becomes a constant, a monotonic function, or a function contains only a single extrema, from which no more IMF can be extracted. Even for data with zero mean, the final residue still could be different from zero. If the data have a trend, the final residue should be that trend. By summing up Equations (16) and (17), we finally obtain

$$X(t) = \sum_{j=1}^{n} c_j + r_n . \tag{2.8}$$

Thus, sifting process produces a decomposition of the data into n-intrinsic modes, and a residue, r_n. When apply the EMD method, a mean or zero reference is not required; EMD only needs only the locations of the local extrema. The sifting process generates the zero reference for each component. Without the need of the zero reference, EMD avoids the troublesome step of removing the mean values for the large non-zero mean.

Two special notes here deserve our attention. First, the sifting process offered a way to circumvent the difficulty of define the local mean in a nonstationary time series, where no length scale exists for one to implement the traditional mean operation. The envelope mean employed here does not involve time scale; however, it is local. Second, the sifting process is a Reynolds-type decomposition: separating variations from the mean, except that the mean is a local instantaneous

mean, so that the different modes are almost orthogonal to each other, except for the nonlinearity in the data.

Recent studies by Flandrin et al. (2004) and Wu and Huang (2004) established that the EMD is equivalent to a dyadic filter bank, and it is also equivalent to an adaptive wavelet. Being adaptive, we have avoided the shortcomings of using any *a priori*-defined wavelet basis, and also avoided the spurious harmonics that would have resulted. The components of the EMD are usually physically meaningful, for the characteristic scales are defined by the physical data.

Having established the decomposition, we can also identify a new use of the IMF components as filtering. Traditionally, filtering is carried out in frequency space only. But there is a great difficult in applying the frequency filtering when the data is either nonlinear or nonstationary or both, for both nonlinear and nonstationary data generate harmonics of all ranges. Therefore, any filtering will eliminate some of the harmonics, which will cause deformation of the data filtered. Using IMF, however, we can devise a time space filtering. For example, a low pass filtered results of a signal having n-IMF components can be simply expressed as

$$X_{lk}\ (t)\ =\ \sum_{k}^{n} c_j\ +\ r_n\ ; \tag{2.9}$$

a high pass results can be expressed as

$$X_{hk}\ (t)\ =\ \sum_{1}^{k} c_j\ ; \tag{2.10}$$

and a band pass result can be expressed as

$$X_{bk}\ (t)\ =\ \sum_{b}^{k} c_j\ . \tag{2.11}$$

The advantage of this time space filtering is that the results preserve the full nonlinearity and nonstationarity in the physical space.

Having obtained the Intrinsic Mode Function components, we can compute the instantaneous frequency for each IMF component as the derivative of the phase function. And we can also designate the instantaneous amplitude from the Hilbert transform to each IMF component. Finally, the original data can be expressed as the real part, RP, of the sum of the data in terms of time, frequency and energy as:

$$X(t)\ =\ RP\sum_{j=1}^{n}\ a_j(t)\ e^{i\ \int\ \omega_j(t)\ dt}\ . \tag{2.12}$$

Equation (22) gives both amplitude and frequency of each component as a function of time. The same data, if expanded in a Fourier representation, would

have a constant amplitude and frequency for each component. The contrast between EMD and Fourier decomposition is clear: The IMF represents a generalized Fourier expansion with a time varying function for amplitude and frequency. This frequency-time distribution of the amplitude is designated as the Hilbert Amplitude Spectrum, *H(s, t)*, or simply the Hilbert spectrum.

From the Hilbert spectrum, we can also define the marginal spectrum, *h(sn)*, as

$$h(\omega) = \int_0^T H(\omega, t)dt. \tag{2.13}$$

The marginal spectrum offers a measure of total amplitude (or energy) contribution from each frequency value. It represents the cumulated amplitude over the entire data span in a probabilistic sense.

The combination of the Empirical Mode Decomposition and the Hilbert Spectral Analysis is designated by NASA as the Hilbert-Huang Transform (HHT) for short. Recent studies by various investigators indicate that HHT is a super tool for time-frequency analysis of nonlinear and nonstationary data (Huang and Attoh-Okine, 2005, Huang and Shen, 2005). It is based on an adaptive basis, and the frequency is defined through the Hilbert transform. Consequently, there is no need for the spurious harmonics to represent nonlinear waveform deformations as in any of the *a priori* basis methods, and there is no uncertainty principle limitation on time or frequency resolution from the convolution pairs based also on *a priori* bases. A summary of the comparison between Fourier, Wavelet and HHT analyses is given in Table 1.

After this basic development of the HHT method, there are some recent developments, which have either added insight to the results or enhanced the statistical significance of the results. Some of the recent developments are summarized in the following section.

3. An Alternative View on Nonlinearity

Having presented the Hilbert spectral analysis, we will explore the alternative view of Hilbert analysis on nonlinearity effects in the data. When one decomposing any data with an *a priori* basis, an inevitable consequence is to have harmonics, which are mathematic artifacts rather than physical entities. Take the water surface waves as an example, which are certainly nonlinear. Therefore, in the traditional view, we have to employee harmonics of the fundamental to fit the nonlinearly distorted profile. Yet, all of the harmonics are not dispersive; they are all bounded waves and have to propagate at the same phase speed as the fundamental. As a result, the wave spectra of water waves based on Fourier analysis is an entangled and inseparable mixture of bounded and free waves. Thus it makes the interpretation of the spectrum extremely difficult for any range other than the energy containing part (see, Huang, et al., 1998, 1999). The intra-wave modulation through Hilbert spectral analysis offers a physically meaningful alternative. A simple example as given by Huang et al (1998) is the mathematic model,

TABLE 1. Comparisons between Fourier, Wavelet and Hilbert-Huang Transform in Data analysis.

	Fourier	Wavelet	Hilbert
Basis	*a priori*	*a priori*	Adaptive
Frequency	Convolution: Global, Uncertainty	Convolution: Global, Uncertainty	Differentiation: Local, Certainty
Presentation	Energy-frequency	Energy-time-frequency	Energy-time-frequency
Nonlinear	No	No	Yes
Non-stationary	No	Yes	Yes
Feature extraction	No	Discrete: no Continuous: yes	Yes
Theoretical base	Theory	Theory	Empirical

$$x(t) = \cos(\alpha t + \varepsilon \sin 2\alpha t), \tag{3.1}$$

which has an intra-wave modulated instantaneous frequency of

$$\omega(t) = \alpha\,(1 + 2\varepsilon \cos \alpha t\,). \tag{3.2}$$

This frequency truthfully depicts the behavior of the oscillator. Yet using Fourier representation the data would have to be decomposed into the fundamental and harmonics as

$$x(t) = \left(1 - \frac{\varepsilon}{2}\right) \cos \alpha t + \frac{\varepsilon}{2} \cos 3\alpha t + \ldots. \tag{3.3}$$

Although the two representations are equally valid mathematically, the intra-wave approach is obviously more physically meaning. For more complicated cases, examples can be found in Huang et al (1998, 1999).

4. The Recent Developments

After considering the basics of HHT analysis, some recent developments in the following areas will be discussed in some details:

4.1 The Normalized Hilbert Transform

4.2 Confidence Limit

4.3 Statistical Significance of IMFs

4.1. The Normalized Hilbert Transform

It is well known that, although the Hilbert transform exists for any function of *Lp* class, the phase function of the transformed function will not always yield physically meaningful instantaneous frequencies. The limitations have been summarized succinctly in two theorems:

First, in order to separate the contribution of the phase variation into the phase and amplitude parts, the function have to satisfy the limitation stipulated in the Bedrosian theorem (1963), which states that the Hilbert transform for the product of two functions, *f(t)* and *h(t)*, can be written as

$$H\,[f(t)\,h(t)\,]\;=\;f(t)\,H\,[\,h(t)\,]\;, \tag{4.1}$$

only if the Fourier spectra for *f(t)* and *h(t)* are totally disjoint in frequency space, and the frequency content of the spectrum for *h(t)* is higher than that of *f(t)*. This limitation is critical, for we need to have

$$H[\,a(t)\,\cos\,\theta(t)\,]\;=\;a(t)\;H\,[\cos\,\theta(t)\,]\;, \tag{4.2}$$

otherwise, we cannot use Equation (6) to define the phase function, for the amplitude variation would mix with the phase function. Bedrosian theorem requires that the amplitude is varying be so slowly that the frequency spectra of the envelope and the carrier waves are disjoint. This is possible only for trivial cases, for unless the amplitude is constant, any local deviation can be considered as a sum of delta-functions, which has a wide white spectrum. Therefore, the spectrum for varying amplitude would never be totally separate from that of the carrier. This limitation has made the application of the Hilbert transform even to IMFs problematic. To satisfy this requirement, Huang and Long (2003) have proposed the normalization of the IMFs in the following steps: Starting from an IMF, one first finds all the maxima of the IMFs, defining the envelope by spline through all the maxima, and designating the envelope as *E(t)*. Now, normalize the IMF by dividing the IMF by *E(t)*. Thus, we have the normalized function having amplitude always equal to unity. Thus we have circumvented the limitation of Bedrosian theorem.

Then, there is the new restriction given by the Nuttall theorem (1966), which stipulates that the Hilbert transform of cosine is not necessarily the sine with the same phase function for a cosine with an arbitrary phase function. Nuttall gave an energy based error bound, *E*, defined as the difference between *y(t)*, the Hilbert transform of the data, and *Q(t)*, the quadrature (with phase shift of exactly 90a) of the function as

$$\Delta E\;=\;\int\limits_{t=0}^{T}|y(t)\;-\;Q(t)\,|^2\;dt\;=\;\int\limits_{-\infty}^{0}S_q(\omega)\,d\omega\;, \tag{4.3}$$

in which S_q is Fourier spectrum of the quadrature function. Though the proof of this theorem is rigorous, the result is hardly useful, for it gives a constant error

bound over the whole data range. With the normalized IMF, Huang and Long (2003) have proposed a variable error bound based on a simple argument, which goes as follows: compute the difference between squared amplitude of the normalized IMF and unity. If the Hilbert transform is exactly the quadrature, the difference between it and unity should be zero; otherwise, the Hilbert transform cannot be exactly the quadrature. Consequently, the error can be measured simply by the difference between the squared normalized IMF and unity, which is a function of time. Huang and Long (2003) and Huang et al. (2005) have conducted detailed comparisons and found the result quite satisfactory.

Even with the error indicator, we can only know that the Hilbert transform is not exactly the quadrature; we still do not have the correct answer. This prompts a drastic alternative, eschewing the Hilbert transform totally. An exact quadrature has been found (Huang et al., 2005), and it would resolve the difficulties associated with the instantaneous frequency computation.

4.2. The Confidence Limit

The confidence limit for the Fourier spectral analysis is based on the ergodic theory, where the temporal average is treated as the ensemble average. This approach is only valid if the processes are stationary. Huang et al. (2003) has proposed a different approach by utilizing the fact that there are infinite many ways to decompose one given function into difference components. Using EMD, we can still obtain many different sets of IMFs by changing the stoppage criteria by changing the *S*-number. The confidence limit so derived does not depend on the ergodic theory.

From the confidence limit study, Huang et al. (2003) also found the optimal *S*-number, when the differences reach a local minimum. Based on their limited experience from different data sets, they concluded that an *S*-number in the range of 4 to 8 performed well. Logic also dictates that the *S*-number should not be too high (which would drain all the physical meaning out of the IMF), nor too low (which would leave some riding waves remaining in the resulting IMFs).

4.3. The Statistical Significance of IMFs

The EMD is a method to separate the data into different components by their scales. There is always the question: On what is the statistical significance of the IMFs based? In data containing noise, how can we separate the noise from information with confidence? This question was addressed by both Flandrin et al. (2004) and Wu and Huang (2004) through the study of signals consisting of noise only. Using white noise, Wu and Huang (2004) found the relationship between the mean period and RMS values of the IMFs. Furthermore, from the statistical properties of the scattering of the data, they found the bounds of the data distribution analytically. They concluded that when a data set is analyzed with EMD, if the mean period-RMS values exist within the noise bounds, the components most likely represent noise. On the other hand, if the mean period-RMS values exceed the noise bounds, then those IMFs must represent statistically significant information.

5. Mathematical Problem Associated with HHT

HHT is an empirically based method. This limitation is not severe when we consider it as a data analysis tool, for all the data are empirical values without analytic expressions anyway. We are at the stage of the wavelet analysis in the earlier 80s: producing useful results but waiting for mathematical foundation to rest our case. The outstanding mathematical problems, as listed by Huang (2005), are summarized here. We hope the mathematicians working in wavelet analysis will be interested in this new alternative and help as follows:

A. Adaptive data analysis methodology in general
B. Nonlinear system identification methods
C. Prediction problem for nonstationary processes (end effect)
D. Spline problem (best spline implement of HHT, convergence and 2-D)
E. Optimization problem (the best IMF selection and uniqueness)
F. Approximation problem (Hilbert transform and quadrature)
G. Miscellaneous questions concerning the HHT

6. Conclusions

The combination of EMD and HSA has provided an adaptive method to analyze nonstationary and nonlinear time series. It can perform and enhance most of the traditional data analysis tasks, such as filtering, regressions, and spectral analysis adaptively. Although adaptive signal analysis is long sought goal for the engineering community (Windrows and Stearns, 1985), the requirement here is much more stringent: we have to deal both nonlinearity and nonstationarity; therefore, the simple feedback method used for stationary processes would not be sufficient. This stringent requirement has put the new method on an empirical base at the present time. As far as data analysis is concerned, the lack of analytic expression would not be a problem, for none of the data came in analytical form anyway. Nevertheless, a purely empirical approach will certainly present a problem for a rigorous mathematical proof of the validity of the method. It is an earnest hope that the usefulness of the method will eventually interested the mathematicians to examine the method critically and constructively, so that the method will find its mathematical foundation established rigorously similar to what Daubechies (1992) had done for the Wavelet analysis.

References

[1] Bedrosian, E., 1963: *On the quadrature approximation to the Hilbert transform of modulated signals,* Proc. IEEE, 51, 868-869. Cohen, L., 1995: Time-frequency Analysis, Prentice Hall, Englewood Cliffs, NJ

[2] Diks, C., 1997: *Nonlinear Time Series Analysis,* World Scientific Press, Singapore.

[3] Daubechies, I., 1992: *Ten Lectures on Wavelets,* Philadelphia SIAM.

[4] Flandrin, P. 1999: *Time-Frequency / Time-Scale Analysis.* Academic Press, San Diego, CA.

[5] Flandrin, P., Rilling, G. and Gonçalvès, P., 2004: *Empirical mode decomposition as a filterbank.* IEEE Signal Proc Lett. 11 (2): 112-114.

[6] Hahn, S. L. 1996: *Hilbert Transforms in Signal Processing.* Artech House, Boston, MA

[7] Huang, N. E., 2005: *Introduction to Hilbert-Huang transform and its associated mathematical problems. 1-32, Hilbert-Huang Transform in Engineering,* Ed. N. E. Huang and N. Attoh-Okine, 2005, CRC Press, New York.

[8] Huang N. E., S. R. Long, and Z. Shen, 1996: *Frequency Downshift in Nonlinear Water Wave Evolution.* Advances in Appl. Mech. 32, 59-117.

[9] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, S. H., Zheng, Q., Tung, C. C. and Liu, H. H. 1998 *The empirical mode decomposition method and the Hilbert spectrum for non-stationary time series analysis,* Proc. Roy. Soc. London, A454, 903-995.

[10] Huang, N. E., Z. Shen, R. S. Long, 1999: *A New View of Nonlinear Water Waves - The Hilbert Spectrum,* Ann. Rev. Fluid Mech. 31, 417-457.

[11] Huang, N. E., Wu, M. L., Long, S. R., Shen, S. S. P., Qu, W. D., Gloersen, P. and Fan, K. L. 2003 *A confidence limit for the empirical mode decomposition and the Hilbert spectral analysis,* Proc. of Roy. Soc. London, A459, 2317-2345.

[12] Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Blank, K., Liu, T. W. 2005 *On instantaneous frequency,* Proc. of Roy. Soc. London (Submitted)

[13] Huang, N. E. and Long, S. R. 2003 *A generalized zero-crossing for local frequency determination.* US Patent pending.

[14] Huang, N. E. and N. Attoh-Okine, Ed. 2005: *Hilbert-Huang Transform in Engineering,* CRC Press, New York.

[15] Huang, N. E. and Samuel S. S. Shen, Ed. 2005: *Introduction to Hilbert-Huang Transform and Applications.* 311 pp, World Scientific, Singapore.

[16] Kantz, H. and T. Schreiber, 1997: *Nonlinear Time Series Analysis,* Cambridge University Press, Cambridge.

[17] Melville, W. K., 1983: *Wave modulation and breakdown.* J. Fluid Mech. 128, 489-506.

[18] Nuttall, A. H., 1966: *On the quadrature approximation to the Hilbert Transform of modulated signals,* Proceedings of IEEE, 54, 1458-1459.

[19] Tong, H., 1990: *Nonlinear Time Series Analysis,* Oxford University Press, Oxford.

[20] Windrows, B and S. D. Stearns, 1985: *Adaptive Signal Processing,* Prentice Hall, Upper Saddle River, NJ

[21] Wu, Z. and Huang, N. E. 2004: *A study of the characteristics of white noise using the empirical mode decomposition method*, Proc. Roy. Soc. London, A460, 1597-1611.

Norden E. Huang
Research Center for Data Analysis
National Central University
Chungli, Taiwan 32001 ROC
e-mail: Norden@ncu.edu.tw