Structural damage detection and estimation by amplitude and frequency modulation analysis

P. Frank Pai
Department of Mechanical and Aerospace Engineering
University of Missouri, Columbia, MO USA 65211

ABSTRACT

Presented here is an amplitude and frequency modulation method (AFMM) for extracting damage-induced nonlinear characteristics and intermittent transient responses by processing steady-state/transient responses using the empirical mode decomposition, Hilbert-Huang transform (HHT), and nonlinear dynamic characteristics derived from perturbation analysis. A sliding-window fitting (SWF) method is derived to show the physical implication of the proposed method and other methods for time-frequency signal decomposition. Similar to the short-time Fourier transform and wavelet transform the SWF uses windowed regular harmonics and function orthogonality to extract time-localized regular and/or distorted harmonics. On the other hand the HHT uses the apparent time scales revealed by the signal's local maxima and minima to sequentially sift components of different time scales, starting from high-frequency to low-frequency ones. Because HHT does not use predetermined basis functions and function orthogonality for component extraction, it provides more accurate instant amplitudes and frequencies of extracted components for accurate estimation of system characteristics and nonlinearities. Moreover, because the first component extracted from HHT contains all original discontinuities, its time-varying amplitude and frequency are excellent indicators for pinpointing times and locations of impulsive external loads and damages that cause intermittent responses. However, the discontinuity-induced Gibbs' effect makes HHT analysis inaccurate around the two data ends. On the other hand, the SWF analysis is not affected by Gibbs' effect, but it cannot extract accurate time-varying frequencies and amplitudes. Numerical results show that the proposed AFMM can provide accurate estimations of softening and hardening effects, different orders of nonlinearity, linear and nonlinear system parameters, and time instants of intermittent transient responses for damage detection and estimation.

Keywords: Identification of damage-induced nonlinearities, amplitude and frequency modulations, signal decomposition, Hilbert-Huang transform, perturbation analysis, damage estimation

1. INTRODUCTION

Dynamics of damaged structures is essentially nonlinear [1-7], and structures are often built with nonlinearities and/or forced to vibrate nonlinearly under normal operation conditions [7-10]. Nonlinearities can be due to large elastic deformations (i.e., geometric nonlinearity), deformation-dependent material properties (i.e., material nonlinearity), backlash, clearances between mounting brackets, geometric constraints on deformation, misalignment of substructures, dry friction, and many types of nonlinear hysteretic damping (e.g., aerodynamic damping, damping of shock absorbers for vehicles, and material damping of shape memory alloys and other materials). Damage introduces extra nonlinearities and aging aggravates nonlinearities [1,2,4,7]. For example, dry friction between fractured surfaces is a nonlinear effect, and the breathing of a crack causes nonlinear intermittent transient response. Also, demountable/retractable civil structures are built with nonlinearities caused by loose joints, and the usage/aging aggravates such nonlinearities. Hence, health monitoring and control of structures highly depends on accurate nonlinearity identification, which is why industries, DOD agencies, and NASA centers are always looking for advanced nonlinearity identification techniques.

Health monitoring and nonlinearity identification is challenging reverse engineering. For several decades structural engineers have been developing dynamics-based methods for rapid damage inspection of large structures [1-4]. Based on the complexity of sensor systems, signal processing methods, and accuracy of derived damage indicators, dynamics-based damage detection methods can be separated into three groups. Methods in the first group require a simultaneous full-field measurement method/tool (e.g., Moire interferometry, digital shearography, and scanning laser vibrometers), and they process the measured displacement, slope, or velocity field to compute curvatures and/or strains and then locate damage by examining abnormality or sudden change of these spatially distributed data [6,11,12].
Methods in the second group require simultaneous measurements of many points, and a well-calibrated structural model and a modal expansion/update method are used to locate damage [4,13]. Methods in the third group require simultaneous measurements of only a few locations, process the measured time traces to extract dynamic characteristics (i.e., natural frequencies, damping ratios, wave propagation characters, and nonlinear effects) to reveal the existence of damage, and use the traveling sequence of abnormality to locate damage [1,14-16]. The third approach with time-of-flight analysis is commonly used for in-work damage detection, and it is more practical and economical than other methods for health monitoring. However, the challenges are how to extract dynamic characteristics from nonlinear noise-contaminated signals and how to correlate the responses measured at just a few physical locations to deduce unique damage indicators. Moreover, it is important to design experiments to limit the number of parameters to be extracted from each experiment in order to increase the probability of unique solutions. It is even more important to have a signal processing and data mining method that can extract from each set of experimental data as many system parameters as possible in order to reduce the time and cost of experiments. Hence, advanced nonlinear signal processing plays the key role in nonlinearity identification and health monitoring of structures.

The paper presents a time-frequency signal decomposition and amplitude and frequency modulation technique for nonlinearity identification and health monitoring of structural systems.

2. LIMITATIONS OF AVAILABLE SIGNAL PROCESSING METHODS

2.1 Nonlinear dynamic characteristics from perturbation analysis

To reveal limitations of available signal processing methods and to illustrate derivation of formulas for nonlinearity identification we consider the following Duffing oscillator and the corresponding second-order perturbation solution:

\[ \ddot{u} + 2\zeta \omega_0 u + \omega_0^2 u + \alpha u^3 = F \cos \Omega t \quad (\sigma = \Omega - \omega) \quad (1a) \]

\[ u(t) = a \cos(\Omega t - \phi) + a_3 \cos(3\Omega t - 3\phi) = \dot{a} \cos(\Omega t - \Theta) \quad \left( a \equiv a_0 \equiv \frac{\alpha a^3}{32\Omega^2} \right) \quad (1b) \]

\[ \dot{\hat{a}}(t) = \sqrt{a^2 + a_3^2 + 2aa_3 \cos(2\Omega t - 2\phi)} \approx a + a_3 \cos(2\Omega t - 2\phi) \quad (1c) \]

\[ \Theta(t) = \tan^{-1} \frac{a_3 \sin(2\Omega t - 2\phi)}{a + a_3 \cos(2\Omega t - 2\phi)} \approx \frac{a_3}{a} \sin(2\Omega t - 2\phi) \quad (1d) \]

\[ \dot{\Omega}(t) = \Omega + \Theta \approx \Omega + \frac{2\Omega a_3}{a} \cos(2\Omega t - 2\phi), \quad \frac{\Omega}{a} \frac{d\dot{a}}{dt} = \frac{\dot{\Omega}}{a} \approx 0.5 \quad (1e),(1f) \]

\[ \left( \frac{3\alpha}{8\omega} \right)^2 \dot{a}^2 - \frac{3\alpha \sigma}{4\omega} a_4 + \left( \sigma^2 + \sigma_0^2 \right) \dot{a}^2 - \left( \frac{F}{2\omega} \right)^2 = 0 \quad (1g) \]

\[ \phi = \tan^{-1} \frac{\omega_0}{3\alpha a^2 (8\omega) - \sigma}, \quad \dot{\omega}_0 = \omega + \frac{3\alpha a^2}{8\omega} \quad (1h),(1i) \]

where \( \zeta, \omega, \alpha, F, a, a_3, \) and \( \phi \) are constants. The amplitude \( a \), phase angle \( \phi \), and undamped natural frequency \( \dot{\omega}_0 \) are determined by (1g)-(1i), which are derived from perturbation analysis [17]. Eq. (1b) shows that \( u(t) \) consists of two synchronous harmonics (i.e., reaching local maxima at the same time), but, because \( \ddot{a} \equiv a_0 \), \( u(t) \) appears in simulation as one distorted harmonic having its amplitude \( \hat{a} \) and frequency \( \dot{\Omega} \) varying at a frequency \( 2\Omega \). This phenomenon and \( (\Omega/a)(d\dot{a}/d\dot{\Omega}) \approx 0.5 \) can be used to determine the order (cubic or other) of nonlinearity. Moreover, if \( \alpha > 0 \), \( a_3 / a > 0 \) and \( \dot{\Omega} \) and \( \hat{a} \) are at their local maxima when \( u(t) \) is at its local maxima or minima. If \( \alpha < 0 \), \( a_3 / a < 0 \) and \( \dot{\Omega} \) and \( \hat{a} \) are at their local minima when \( u(t) \) is at its local maxima or minima. This phenomenon can be used to determine the type (hardening or softening) of nonlinearity. Furthermore, the magnitude of nonlinearity can be estimated using the average amplitude \( a \) and the variation amplitude \( a_3 \) of \( \hat{a}(t) \) as \( \alpha = 32\Omega^2 a_3 / a^3 \). Hence, a distorted harmonic itself is actually more useful than two separate, synchronous regular harmonics for nonlinearity identification using time-domain data.
On the other hand, it is difficult to identify nonlinearity using frequency-domain data. Because multiple steady-state solutions coexist for one \( \Omega \), nonlinear frequency response curves cannot be experimentally obtained using random excitations, impulse excitations, chirp excitations, or other broad-band excitations [18,19]. A broad-band excitation simultaneously excites several modes of a continuous system. If each modal vibration is a distorted harmonic wave due to nonlinearity, the frequency response curve loses its physical meaning because each distorted harmonic wave is represented by several spectral lines, and peaks due to higher harmonics (see (1b)) may be erroneously identified as modes. More seriously, it is difficult to determine the order and magnitude of nonlinearity from frequency response curves because the backbone curves (i.e., (1i)) cannot be determined since the unstable branch cannot be experimentally obtained. Furthermore, a frequency response curve cannot reveal transient nonlinearities because it is an averaged presentation of the whole sampled period.

Because identification of stiffness and damping in time domain requires free damped or forced transient response, how to accurately extract time-varying dynamic characteristics from non-stationary response is important but challenging because, except the distorted harmonic response (i.e., intrawave amplitude and frequency modulations) to a harmonic excitation, many other nonlinear phenomena may coexist. For example, a nonlinear system has amplitude-dependent natural frequencies (see (1i)), and it may vibrate at a frequency much lower than the externally applied harmonic excitation [20]. Other nonlinear phenomena include multiple-mode vibrations caused by external resonances and/or internal resonances (i.e., interwave modulation), intermittent nonlinear response, bifurcation, and chaotic vibrations [10,21]. Hence, nonlinearity identification and health monitoring requires a signal processing technique that can extract accurate time-varying dynamic characteristics from nonlinear non-stationary responses and pinpoint time instants of intermittent transient responses.

### 2.2 Problems of available time-frequency decomposition methods

For an arbitrary time signal \( u(t) \), the discrete Fourier transform (DFT) decomposes it into [22]

\[
\begin{align*}
    u(t_k) &\equiv u_k = a_0 + 2 \sum_{i=1}^{N/2} \left( a_i \cos \omega_k t_k + b_i \sin \omega_k t_k \right) = a_0 + \text{Real} \left( 2 \sum_{i=1}^{N/2} U_i e^{j \omega_k t_k} \right) \\
    a_i &= \frac{1}{N} \sum_{k=1}^{N} u_k \cos \omega_k t_k, \quad b_i = \frac{1}{N} \sum_{k=1}^{N} u_k \sin \omega_k t_k, \quad U(\omega) \equiv U_i = a_i - j b_i = \frac{1}{N} \sum_{k=1}^{N} u_k e^{-j \omega_k t_k} \tag{2}
\end{align*}
\]

where \( \omega_k = 2 \pi k / T \), \( j = \sqrt{-1} \), \( N \) is the total number of samples, \( t_k = k \Delta t \), \( \Delta t \) is the sampling interval, \( T (= N \Delta t) \) is the sampled period, the Nyquist frequency is \( 1/(2 \Delta t) \), and \( U_i \) is the Fourier spectrum. Eq. (2) shows that orthogonality between \( u(t) \) and \( \cos \omega t \) and \( \sin \omega t \) is used to extract constant-amplitude regular harmonics from \( u(t) \).

If \( v(t) \) denotes the Hilbert transform (HT) of \( u(t) \) (i.e., \( v = H[u] \)), it can be shown that [23,24]

\[
\begin{align*}
    v(t_k) &\equiv v_k = 2 \sum_{i=1}^{N/2} \left( a_i \sin \omega_k t_k - b_i \cos \omega_k t_k \right) = \text{Imag} \left( 2 \sum_{i=1}^{N/2} U_i e^{j \omega_k t_k} \right) \\
    V(\omega) &\equiv V_i = -b_i - j a_i = -j U_i = \frac{1}{N} \sum_{k=1}^{N} u_k e^{-j (\omega_k t_k + \pi/2)} \tag{3}
\end{align*}
\]

In other words, \( v(t) \) is obtained by shifting each harmonic of \( u(t) \) by \(-90^\circ\). Moreover, (2) shows that the Fourier spectrum needed for computing \( v(t) \) can be used to compute the instantaneous velocity and acceleration by using the inverse discrete Fourier transform (IDFT) as

\[
\begin{align*}
    u(t) + j v(t) &= a_0 + 2 \sum_{i=1}^{N/2} U_i e^{j \omega t}, \quad \dot{u}(t) + j \ddot{v}(t) = 2 \sum_{i=1}^{N/2} j \omega_i U_i e^{j \omega t}, \quad \ddot{u}(t) + j \dddot{v}(t) = -2 \sum_{i=1}^{N/2} \omega_i^2 U_i e^{j \omega t} \tag{4}
\end{align*}
\]

Unfortunately, if \( u(0) \neq u(T) \), \( U_i \) will contain many high-frequency harmonics that result in leakage in frequency domain and Gibbs' effect in time domain [22,23,25]. Leakage makes it difficult to understand the Fourier spectrum. Gibbs' effect happens because continuous functions (i.e., \( \cos \omega t \) and \( \sin \omega t \) in (2)) are used to fit a discontinuous function [23,26].

Fourier transform (FT) is often used to transform time-domain data into frequency-domain data to identify system characteristics from the obtained spectrum, but a spectrum cannot show time-varying characteristics of the processed signal [18,19]. To overcome this problem, short-time Fourier transform (STFT, spectrogram), wavelet transform (WT), and Hilbert-Huang transform (HHT) have been developed for time-frequency analysis. STFT uses windowed regular harmonics and function orthogonality to simultaneously extract time-localized regular harmonics, and

Proc. of SPIE Vol. 6935  69351S-3
WT uses scaled and shifted wavelets and function orthogonality to simultaneously extract time-localized components similar to the wavelets [27,28]. Unfortunately, STFT and WT cannot accurately extract time-varying frequencies and amplitudes because they use pre-determined basis functions and function orthogonality to simultaneously extract components [29,30]. On the other hand, HHT uses the apparent time scales revealed by the signal's local maxima and minima to sequentially sift components of different time scales, starting from high-frequency to low-frequency ones [31-35]. Because HHT does not use predetermined basis functions and function orthogonality for component extraction, it provides accurate time-varying amplitudes and frequencies of extracted components [36-38].

HHT combines the empirical mode decomposition (EMD) and Hilbert spectral analysis for accurate time-frequency decomposition of any signals. EMD is a non-causal dyadic filter and is equivalent to an adaptive wavelet [34,39,40]. The first step of HHT is to use EMD to sequentially decompose a time series $u(t)$ into $n$ intrinsic mode functions (IMFs) $c_i(t)$ and a residual $r_n$ as

$$u(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \quad (5)$$

where $c_1$ has the highest frequency and is the first extracted IMF. The characteristic time scale of $c_1$ is defined by the time lapse between the extrema of $u$. Once the extrema are identified, compute the upper envelope by connecting all local maxima using a natural cubic spline, compute the lower envelope by connecting all local minima using another natural cubic spline, subtract the mean of the upper and lower envelopes, $m_1$, from the signal, and then treat the residuary signal as a new signal. Repeat these steps for $K$ times until the left signal has a pair of symmetric envelopes (i.e., $m_K = 0$), and then define $c_1$ as

$$c_1 \equiv u - m_1 - \cdots - m_K \quad (6)$$

This sifting process eliminates low-frequency riding waves and makes the wave profile symmetric. During the sifting process a deviation $D_v$ is computed from the two consecutive sifting results as

$$D_v \equiv \sqrt{\sum_{i=1}^{N} [c_{1k}(t_i) - c_{k-1}(t_i)]^2 / \sum_{i=1}^{N} c_{1k-2}^2(t_i)} \quad (7)$$

where, e.g., $c_{1k} \equiv u - m_1 - \cdots - m_{ik}$, $t_i = i\Delta t$, and $T = N\Delta t$. To end iteration one can limit $D_v$ to a small number and/or limit the maximum number of iterations. After $c_1$ is obtained, define the residual $r_1$, treat $r_1$ as the new data, and repeat the steps shown in (6) to obtain other $c_i$ ($i = 2, \ldots, n$) as

$$c_i \equiv r_{i-1} - m_{1i} - \cdots - m_{ik}, \quad r_{n-1} \equiv u(t) - c_1 - \cdots - c_{n-1} \quad (8)$$

The whole sifting process can be stopped when the residual $r_n$ becomes a monotonic function from which no more IMF can be extracted. For data with a trend, the last IMF $r_n$ should be the trend.

The second step of HHT is to perform Hilbert transform and compute the time-varying frequency $\omega_i$ and amplitude $A_i$ of each $c_i$. After all $c_i(t)$ are extracted, one can perform Hilbert transform to obtain $d_i(t)$ from each $c_i(t)$ and then define $z_i(t)$ and use (5) without $r_n$ to obtain

$$z_i(t) = c_i(t) + jd_i(t) = A_i e^{i\theta_i}, \quad A_i \equiv \sqrt{c_i^2 + d_i^2}, \quad \theta_i \equiv \arctan \frac{d_i}{c_i}$$

$$u(t) = \text{Real} \left( \sum_{i=1}^{n} [c_i(t) + jd_i(t)] \right) = \text{Real} \left( \sum_{i=1}^{n} A_i(t) e^{i\theta_i(t)} \right) \quad (9)$$

To reduce the influence of noise on the calculated frequency $\omega_i$ (equal $d \theta / dt$ at $t = t_*$), each $\omega_i$ can be obtained by averaging over $2\Delta t$ as

$$\omega_i \approx \sum_{m=-\infty}^{\infty} \left[ \theta_i(t_* + m\Delta t) - \theta_i(t_* + (m-1)\Delta t) \right] / (2\Delta t) \quad (10)$$

For damage detection, it is better to compute $\omega_i$ and other time derivatives using
\[ \dot{A}_i = \frac{c_i \dot{d}_i + d_i \ddot{d}_i}{A_i}, \quad \ddot{A}_i = \frac{c_i \ddot{d}_i + d_i \ddot{d}_i - \dddot{A}_i}{A_i}, \quad \omega_i = \frac{c_i \dot{d}_i - \dot{c}_i \ddot{d}_i}{A_i}, \quad \phi_i = \frac{c_i \ddot{d}_i - \ddot{c}_i \dddot{d}_i - 2\omega_i (c_i \dot{c}_i + d_i \ddot{d}_i)}{A_i} \] (11)

which are derived from (9). The \( \dot{c}_i, \ddot{d}_i, \dot{d}_i, \) and \( \dddot{d}_i \) needed in (11) can be computed using the inverse Fourier transform shown in (4) without numerical differentiation in time domain.

Unfortunately, the accuracy of HHT analysis suffers from several mathematical and numerical problems [36-38]. If \( c_i(0) \neq c_i(T) \), Gibbs’ effect reduces the accuracy of frequencies and amplitudes at data ends. Even if the two ends of an extracted IMF is smoothly extended to begin and end with zero, the calculated frequency and amplitude are inaccurate if the envelopes of the two extended segments are asymmetric. To accurately estimate, for example, the \( \alpha \) in (1a), \( \alpha \) needs to be separated from \( \dot{a} \), which HHT cannot do because \( \alpha \) is small. HHT cannot accurately decompose a signal consisting of two non-synchronous regular harmonics with a frequency ratio \( \omega_1 / \omega_2 < 1.5 \) when they reach their local extrema around the same time [37], which causes the serious intermittency problem (i.e., a mode only appears in a segment of time) and decreases the credibility of HHT. If a signal contains two harmonics of close amplitudes and frequencies, the computed time-varying frequency and amplitude are erroneous or inaccurate because its envelope is asymmetric or its amplitude changes sign [37]. If a transient signal contains a fast decaying moving average, the moving average alters the signal’s peak locations and reduces the accuracy of extracted components.

### 2.3 Further Decomposition

The problems of HHT analysis can be overcome by pre-processing the signal before HHT analysis and/or post-processing the extracted IMFs by data extension or further decomposition [35-38]. For further decomposition of IMFs, we propose the following sliding-window fitting (SWF) method. If an IMF \( c(t) \) from HHT analysis contains two major frequencies \( \omega_1 \) and \( \omega_2 \) (e.g., \( \omega_1 = 3\Omega_1, \omega_2 = \Omega_2 \) in (1b)), one can assume that

\[ c(t) = \sum_{k=1}^{2} \left[ e_k \cos(\omega_k t) - \hat{e}_k \sin(\omega_k t) \right] + e_3 + \hat{e}_3 t + e_4 t^2 \]

(12)

where \( e_k \) and \( \hat{e}_k \) are constants, \( \bar{T} (\equiv t - t_i) \) is a moving time coordinate, \( t_i \) is the observed time instant, and

\[ C_k \equiv \sqrt{e_k^2 + \hat{e}_k^2} \cos(\omega_k t_i + \phi_k), \quad \hat{C}_k \equiv \sqrt{e_k^2 + \hat{e}_k^2} \sin(\omega_k t_i + \phi_k), \quad \phi_k \equiv \tan^{-1} \frac{\hat{e}_k}{e_k} \]

(13)

The \( C_k \) and \( \hat{C}_k \) for the data point at \( \bar{T} = 0 \) can be determined using the least-squares fitting by minimizing the square error \( E_{\text{sw}} = \sum_{i=m}^{n} \alpha^H (\hat{c}_i - c_i)^2 \), where \( \hat{c}_i \) denotes the right-hand side of (12) with \( \bar{T} = \bar{T}_i \) and \( c_i \) denotes the actual IMF at \( \bar{T}_i \). The total number of points used is \( 2m+1 \), \( \alpha^H \) is a weighting factor, and the forgetting factor \( \alpha (\leq 1) \) is chosen by the user. After \( C_k \) and \( \hat{C}_k \) are determined, it follows from (12) that

\[ c(t_j) = C_1 + C_2 + C_3, \quad \hat{c}(t_j) = -\omega_1 \hat{C}_1 - \omega_2 \hat{C}_2 + \hat{C}_3, \quad \ddot{c}(t_j) = -\omega_1^2 C_1 - \omega_2^2 C_2 + 2C_4 \]

(14)

It reveals that \( c(t_j) \) consists of the instantaneous value \( C_1 \) of the harmonic \( \cos(\omega_2 \bar{T}) \), the instantaneous value \( C_2 \) of \( \cos(\omega_2 \bar{T}) \), and the low-frequency moving average \( C_4 \). More importantly, each of \( c, \hat{c}, \) and \( \ddot{c} \) is decomposed into three components of different frequencies. Furthermore, it follows from (13) that

\[ A_k \equiv \sqrt{C_k^2 + \hat{C}_k^2} = \sqrt{e_k^2 + \hat{e}_k^2}, \quad \theta_k \equiv \tan^{-1} \frac{\hat{C}_k}{C_k} = \omega_k t_i + \phi_k \]

(15)

where \( A_k \) and \( \hat{A}_k \) are the instantaneous amplitudes of the first and second harmonics, respectively. To reduce the influence of noise on the calculated instantaneous frequency \( \omega_i \), each \( \omega_i \) at \( t = t_i \) can be computed by averaging over \( 2p \Delta t \), as shown in (10). This method can be used to extract as many harmonics as needed by adding to (12) major harmonics.
identified from the signal's Fourier spectrum. Comparing (9) and (15) reveals that \( \hat{C}_k \) is the Hilbert transform of \( C_k \), i.e.,

\[ H(C_k) = \hat{C}_k. \]

The synchronous detection method (SDM) used in radio-signal demodulation and lock-in amplifiers can be extended for further decomposition of an IMF [40]. For example, if the following signal

\[ u(t) = c_1(t) + c_2(t) = A_1(t) \cos(\theta_1(t) + \gamma_1) + A_2(t) \cos(\theta_1(t) + \gamma_2) \]

(16a)
is multiplied by \( \cos(\theta_1(t)) \) and \( \sin(\theta_1(t)) \), respectively, we obtain

\[ \hat{u}(t) = u(t) \cos(\theta_1(t)) = 0.5A_1 \cos(\gamma_1) + 0.5A_1 \cos(2\theta_1 + \gamma_1) + A_2 \cos(\theta_2 + \gamma_2) \cos(\theta_1(t)) \]

\[ \hat{u}(t) = u(t) \sin(\theta_1(t)) = -0.5A_1 \sin(\gamma_1) + 0.5A_1 \sin(2\theta_1 + \gamma_1) + A_2 \cos(\theta_2 + \gamma_2) \sin(\theta_1(t)) \]

(16b)

\[ \sqrt{\hat{u}^2 + \hat{u}^2} = 0.5A_1 + \text{high-frequency components} \]

After the high-frequency components of \( \hat{u} \) and \( \hat{u} \) are filtered out using a low-pass filter, \( A_1(t) \), \( \gamma_1(t) = \tan^{-1}(\hat{u}/\hat{u}) \) and hence \( A_1 \cos(\theta_1 + \gamma_1) \) can be obtained as

\[ A_1 = 2\sqrt{\hat{u}^2 + \hat{u}^2}, \quad A_1 \cos(\theta_1(t) + \gamma_1) = 2[\hat{u}(t) \cos(\theta_1(t)) + \hat{u}(t) \sin(\theta_1(t))] \]

(16c)

A constant \( \omega_0 \) and \( \theta_0(t) = \omega_0 t \) is commonly used in synchronous detection, but one can extend it for cases with time-varying \( \omega_0 \). If the frequency of an IMF is time-varying, one can use the smoothed time-varying frequency (by curve-fitting using low-order polynomials or low-pass filtering) in (16b,c) to extract the main component. The smoothed time-varying frequency can also be used in (12) for the SWF method.

3. NUMERICAL RESULTS

3.1 Nonlinearity identification

We consider the following nonlinear oscillator:

\[ m\ddot{u} + cu + ku + m\alpha_2u^2 + m\alpha_3u^3 + m\beta_2u^2 + m\beta_3\dot{u}^3 \]

\[ = m[\ddot{u} + 2\zeta\omega_0\dot{u} + \omega_0^2u + \alpha_2u^2 + \alpha_3u^3 + \beta_2\dot{u}^2 \text{sign}(\dot{u}) + \beta_3\dot{u}^3] = F_u \cos(\Omega t) \]

(17a)

First we consider the following nonlinear non-stationary vibration:

\[ F_u = 40, \quad \Omega = 1.2\pi, \quad u(0) = 5, \quad \dot{u}(0) = 0, \]

\[ \zeta = 0.005, \quad \omega_0 = 2\pi, \quad m = 2, \quad \alpha_1 = 1, \quad \alpha_2 = \beta_2 = \beta_3 = 0 \]

(17b)

Figs. 1a,b show that \( c_1 \) is the damped natural harmonic and \( c_2 \) is the distorted harmonic caused by the excitation with \( \Omega = 0.6\text{Hz} \) and the cubic nonlinearity. Fig. 1c shows that \( \omega_1 \) modulates at \( 2\omega_1 \) at the beginning, \( \omega_2 \) modulates at \( 1.2\text{Hz} (= 2\Omega) \), and \( \omega_1 \) decreases with \( A_1 \), which indicate the existence of hardening cubic nonlinearity. Moreover, because \( \omega_2 \) is at its maxima when \( c_2 \) is at its maxima or minima, \( \alpha \) is positive. Fig. 1d also shows the existence of cubic nonlinearity because \( A_1 \) modulates at \( 2\omega_1 \) when \( A_1 \) is large. If the \( \omega_1(t) \) and \( A_1(t) \) are curve-fitted using low-order polynomials, the \( \omega_1 - A_1 \) curve shown in Fig. 1e becomes a smooth backbone curve bent to the right indicating a hardening nonlinearity and can be used to estimate \( \alpha \). Without smoothing, the modulation frequency \( 2\omega_1 \) of \( \omega_1(t) \) and \( A_1(t) \) in Figs. 1c,d actually helps the identification of cubic nonlinearity. Because the \( \ln(A_1(t)) \) in Fig. 1f is a straight line, the damping is a linear one. The linear damping ratio \( \zeta \) can be estimated using

\[ \zeta = -d(\ln A_1)/(\omega_1 dt) \approx -[\ln A_1(T) - \ln A_1(0)]/\omega T = 0.0049, \]

where \( T (=60) \) is the sampled period.
For a linear system, the particular solution under a harmonic excitation should be a steady-state response with a constant amplitude starting from the beginning $t=0$. Hence, the gradual increase of $A_2$ to about 0.73 ($<0.778$, the final value) at $t=55$ in Fig. 1d is due to nonlinearity and the coupling of transient and steady-state solutions, and it also explains why it often takes a long time for a nonlinear system to achieve a steady state. Moreover, the average $\omega_1 - A_1$ curve deviates from the perturbation solution in Fig. 1e when $A_1$ is small, indicating that $c_1$ is affected by $c_2$. Hence, to obtain accurate identification, it is better to separately process one transient response and one steady-state response, instead of one response containing both transient and steady-state responses. Another key point here is that it is more accurate to estimate nonlinearities using a large-amplitude transient response than a steady-state response because the latter depends on the use of second-order asymptotic solutions. However, it is easier to detect nonlinearities using a small-amplitude steady-state response than using a small-amplitude transient response. This phenomenon is very useful for early detection of nonlinear structural vibrations (e.g., flutter of aircraft, and chattering of machining) before the vibration is too big to be controlled.

![Fig. 1: HHT analysis of a non-stationary response:](image)

**3.2 Identification of impacts and intermittent transient response**

Because the $m_{lk}$ in (6) is the average of the upper and lower cubic spline envelopes, $m_{lk}$, $\dot{m}_{lk}$ and $\ddot{m}_{lk}$ are continuous time functions and hence the extracted $c_1(t)$ contains all the discontinuities of $u$, $\dot{u}$, and $\ddot{u}$. Hence, the time-varying $\omega_1$ and $A_1$ of $c_1$ become excellent indicators for pinpointing time instants of impacts (or other intermittent transient effects) because impacts introduce discontinuities into $\dot{u}$ and $\ddot{u}$. Moreover, because $c_1(t)$ has the highest frequency and contains all discontinuities of the signal, there are unique methods for extracting discontinuities and filtering high-frequency noise using HHT. For example, one can add a high-frequency harmonic to the signal and then extract it with the discontinuities and/or high-frequency noise inside the signal as $c_1(t)$. After that subtracting the added high-frequency harmonic from $c_1(t)$ yields the discontinuities and/or noise of the original signal.

To show an example, we consider the following Duffing oscillator subjected to a harmonic excitation and five impacts:

$$m(\dddot{u} + 2\zeta\omega\dot{u} + \omega^2u + u^3) = F_0\cos(\Omega t) + p(t)$$

(18a)
\[ p(t) = 3\delta(t - 21) + 3\delta(t - 31) + 2\delta(t - 32) - 3\delta(t - 41) + 3\delta(t - 51) \]

\[ F_0 = 10, \ \Omega = 2\pi, \ u(0) = \dot{u}(0) = 0, \ \zeta = 0.01, \ \omega = 2\pi, \ m = 2 \]

(18b)

where \( \delta(t) \) is the Dirac delta function. Because an impulsive force causes a sudden change of velocity, the forcing function \( p(t) \) causes the velocity to change by \( 3/m, 3/m, 2/m, -3/m, \) and \( 3/m \) at \( t = 21, 31, 32, 41, \) and \( 51 \) seconds, respectively. In the numerical integration, \( N = 2000 \) and \( \Delta t = 60/N \) are used, and a noise \( 0.002 \cdot \text{randn} \) is added to the obtained \( u(t) \). Figs. 2c,d show that the five impact times are clearly shown by the sudden changes of \( \omega_1 \) and \( A_1 \) of the first IMF \( c_1 \) from HHT analysis, and even the residual \( r_1 \) in Fig. 2b also roughly reveal the impact times. When the amplitude increases at the beginning, \( \omega_1 \) increases and modulates at \( 2\omega_1 \), which indicates a hardening cubic nonlinearity. The inaccurate \( \omega_1 \) and \( A_1 \) at the two data ends are caused by Gibbs’ effect. Figs. 2e,f show that the five impact times are also indicated by the sudden changes of \( \omega_1 \) and \( A_1 \) of the only harmonic \( C_1 \) extracted from SWF analysis. Because \( \omega_1 = 1 \text{ Hz} \) is used in (12), when the actual \( \omega_1 \) is much different from \( 1 \text{ Hz} \), the \( \omega_1 \) extracted from SWF analysis shows large variation, as shown in Fig. 2e when \( A_1 \) is large.

For locating the impacts, we also use a high-order wavelet \( \text{db10} \) (Daubechies wavelet in MATLAB) to perform a two-level decomposition (i.e., \( u = \hat{c}_2 + d_1 + d_2 \)). The obtained level-two approximation \( \hat{c}_2 \) is similar to but different from the \( c_1 \) shown in Fig. 2b. Fig. 3 shows that the impact instants are shown by the detail \( d_2 \), but not by the detail \( d_1 \). Although the beginning times of impacts are revealed by the detail \( d_2 \), unlike the \( \omega_1 \) and \( A_1 \) in Figs. 2c-f, the details \( d_1 \) do not have physical meanings at all.

Fig. 2: HHT and SWF analyses of the \( u(t) \) of (18a,b): (a-d) \( u(t), c_1 \& r_1, \omega_1, \) and \( A_1 \) from HHT, and (e,f) \( \omega_1 \) and \( A_1 \) from SWF.
Next we consider (18a,b) with \( p(t) = 0 \), but the numerically integrated solution is changed by -0.1, 0.1, 0.1, 0.05, -0.1, and 0.1 at \( t=11, 21, 31, 32, 41, \) and \( 51 \) seconds, respectively. The disturbances mimic sensor errors. In the numerical integration, \( N = 2000 \) and \( \Delta t = 60/N \) are used, and a noise \( 0.002 \cdot \text{randn} \) is added to the obtained \( u(t) \). Note that it is difficult to locate the 6 discontinuities in Fig. 4a. With the addition of a high-frequency harmonic \( 0.5 \cos(10\pi t) \) to the signal before the EMD process and the subtraction of it from the extracted \( c_1 \), Figs. 4b-d show that the extracted \( c_1 \) is basically the noise and the 6 discontinuities, and the extracted \( c_2 \) is the original solution. Without adding a high-frequency harmonic, Figs. 4e,f show that the 6 discontinuities are revealed by the sudden changes of the \( \omega_1 \) and \( A_1 \) of the first IMF \( c_1 \) because all discontinuities are retained in \( c_1 \). The smooth \( A_2 \) in Fig. 4d confirms that the 6 discontinuities are extracted into \( c_1 \), and a zoom-in view shows that the \( \omega_2 \) in Fig. 4e is almost the same as the \( \omega_1 \) in Fig. 4e without the 6 spikes.

Using the SWF method with \( \omega_1 = 2\pi \) in (12), Figs. 5a,b show that the 6 discontinuities are revealed by the \( \omega_1 \) and \( A_1 \), but the spikes are smaller than those in Figs. 4e,f because they are locally averaged over the window length. Using the SWF method with \( \omega_1 = 10\pi \) and \( \omega_2 = 2\pi \) in (12), Figs. 5c,d show that the 6 discontinuities are also revealed by the \( \omega_2 \) and \( A_2 \) but the spikes are smaller than those in Figs. 5a,b because the discontinuities are shared by the extracted \( C_1 \) and \( C_2 \). However, the \( A_1 \) clearly reveals the 6 discontinuities. The peaks of the \( A_1 \) in Fig. 5d are smaller than those in Fig. 4d and far from the actual discontinuity values because the discontinuities are shared by the extracted \( C_1 \) and \( C_2 \) and are locally averaged. For locating the discontinuities, we also use a high-order Daubechies db10 wavelet to perform a two-level decomposition (i.e., \( u = \hat{c}_2 + d_1 + d_2 \)). Figs. 5e,f show that the 6 discontinuities are revealed by details \( d_1 \) and \( d_2 \), but the spikes are small because the discontinuities are shared by all details.
Fig. 4: HHT analysis of the $u(t)$ of (18a,b) with $p(t) = 0$ and 6 displacement errors: (a-d) $u(t)$, $c_1$ & $c_2$, $\omega_1$ & $\omega_2$, and $A_1$ & $A_2$ (adding a harmonic), and (e,f) $\omega_1$ and $A_1$ (without adding a harmonic).

Fig. 5: SWF and wavelet decomposition of the $u(t)$ of (18a,b) with $p(t) = 0$ and 6 displacement errors: (a,b) $\omega_1$ and $A_1$ from SWF, (c,d) $\omega_2$ and $A_1$ & $A_2$ from SWF with an added frequency, and (e,f) details $d_1$ and $d_2$ from wavelet decomposition.

4. CONCLUDING REMARKS

An amplitude and frequency modulation method based on empirical mode decomposition, Hilbert-Huang transform (HHT), and perturbation analysis is developed for identification of nonlinearities and system parameters by processing transient and steady-state responses. Because HHT does not use function orthogonality for extracting components and it allows the use of distorted harmonics, it can decompose a transient or stationary signal into just a few components and
provide accurate time-varying frequency and amplitude of each component to reveal nonlinear effects. Intrawave amplitude and frequency modulation explains the distortion of harmonics by nonlinear effects. Moreover, because the first component extracted from HHT contains the discontinuities and high-frequency noise in the original signal, its time-varying frequency and amplitude become excellent indicators for pinpointing time instants of impact loadings. Accurate identification of damage-induced nonlinearities is the key for dynamics-based detection and diagnosis of structural damage, and accurate identification of time instants of impacts and intermittent transient response is the key for in-work damage detection and estimation. However, the discontinuity-induced Gibbs' effect at two data ends in HHT analysis needs further study in order to improve its accuracy and robustness, and this amplitude and frequency modulation method needs further experimental validation.

REFERENCES