CIRCULAR INSTANTANEOUS FREQUENCY

P. FRANK PAI

Department of Mechanical and Aerospace Engineering
University of Missouri, Columbia, MO 65211, USA

This work defines the unique instantaneous frequency (IF) of an arbitrary time signal to be the circular instantaneous frequency (cIF) of the curvature radius of the signal’s trajectory on the phase plane, where the signal’s conjugate part is obtained from Hilbert transform (HT). Because a general signal of a dynamical system of multiple degrees of freedom contains multiple modal vibrations, its cIF varies dramatically and is not useful for system identification and other applications. If the signal is decomposed into modal vibration components without moving average, each component has no local extrema within each fundamental period and no local loops on the phase plane, each component’s referred instantaneous frequency (rIF) with respect to the origin on the phase plane may be non-circular but is always non-negative, and the time-varying rIF and referred instantaneous amplitude (rIA) are convenient for combining the use of perturbation analysis for system identification. The empirical mode decomposition (EMD) of Hilbert–Huang transform (HHT) is valuable for decomposing a general nonlinear nonstationary signal into zero-mean intrinsic mode functions (IMFs), and HT enables accurate calculation of rIF and rIA of each IMF. Although the concept of circular frequency cannot be used for signal decomposition, it enables the development of time-domain-only techniques for online frequency tracking. A 5-point frequency tracking method is developed to eliminate the incapability of the original 4-point Teager–Kaiser algorithm (TKA) for frequency tracking of signals with moving averages. Moreover, a 3-point conjugate-pair decomposition (CPD) method is derived based on circle-fitting using a pair of conjugate harmonic functions. It is shown that both CPD and TKA are based on the concept of circle fitting, but TKA uses finite difference and CPD uses curve fitting in numerical implementation. However, the accuracy of TKA is easily destroyed by noise because of the use of finite difference. On the other hand, because CPD is based on curve fitting, noise filtering is an implicit capability and its accuracy increases with the number of processed data points. The rIF from HHT and the cIF from CPD and TKA are different by definition. Moreover, because the instantaneous frequency and amplitude are assumed to be constant in CPD and TKA, the cIF from CPD and TKA also deviates from the exact cIF.

Keywords: Circular instantaneous frequency, time–frequency analysis, conjugate-pair decomposition, Teager–Kaiser algorithm, Hilbert–Huang transform.

1. Introduction

Signal processing plays the key role in dynamics-based damage inspection and health monitoring of structures and mechanical systems, and the focus is always on accurate extraction/tracking of dynamic characteristics in order to identify
system parameters (e.g. natural frequencies, mode shapes, mass, damping, stiffness, and external loading) by reverse engineering [Doebling et al. (1996); Sohn et al. (2003)]. Unfortunately, dynamic responses of damaged structures and mechanical systems are often nonlinear and nonstationary because their frequencies and amplitudes change with time [Doebling et al. (1996); Sohn et al. (2003); Huang and Shen (2005)]. Hence, *time-frequency analysis* becomes necessary for extracting/tracking the time-varying instantaneous frequency (IF) and instantaneous amplitude (IA) of such amplitude-modulated and frequency-modulated (AM-FM) signals [Huang and Shen (2005)].

The concept of IF is also important for many other engineering applications. For example, a frequency-modulated (FM) chirp excitation function $a \cos \theta$ with $a$ being constant and $\dot{\theta}(\equiv d\theta/dt)$ being a linearly or nonlinearly varying frequency can be used to quickly obtain a mechanical system’s experimental frequency response function. In communication, an FM signal can be used to transfer a signal $m(t)(|m| \leq 1)$ by embedding it into $a \cos \theta$ such that $\dot{\theta} = \omega_c + m\omega_d$, where $\omega_c$ and $\omega_d(\ll \omega_c)$ are the pre-determined constant carrier and deviation frequencies, respectively. For such an FM signal $a \cos \theta(\equiv u)$ having a constant amplitude, $\dot{\theta}$ is easily accepted to be the signal’s IF without any doubts and is called a *circular instantaneous frequency* (cIF) in this work. (We call it “circular instantaneous frequency” instead of “instantaneous circular frequency” because “instantaneous frequency” and its acronym IF have been well accepted in the literature.) The reason is that, if $a \sin \theta(\equiv v)$ is the signal’s conjugate part, the signal’s trajectory on the $u-v$ (or phase) plane is a circle centered at the origin $(u,v) = (0,0)$. On the other hand, the IF and IA of an AM-FM signal are not well defined in the literature because the trajectory is not circular. For a non-circular trajectory, the frequency referred to the origin at $(u,v) = (0,0)$ is non-circular and will be called a *referred instantaneous frequency* (rIF).

Because time-varying IF and IA provide better details for understanding and identification of nonlinear systems, it is important to have accurate extraction of them by time-frequency analysis. Unfortunately, appropriate time–frequency analysis of nonlinear nonstationary signals is one problem [Huang and Shen (2005)]. Wavelet transform prevails in the analysis of nonstationary signals, but it is not appropriate for nonlinear signals because pre-determined basis functions (i.e. wavelets) are used. For a nonlinear nonstationary signal, the convolution computation of the pre-determined basis functions and the signal creates more problems than solutions because of the time-varying characteristics of a nonlinear signal [Huang and Shen (2005)]. Another serious problem in time–frequency analysis is that the definition, uniqueness, physical meaning, and computation of IF and IA are not clear even up to this moment in the literature [Huang et al. (2009); Huang and Attoh-Okine (2005); Gröchenig (2001); Kumaresan and Rao (1999); Flandrin (1999); Longhin and Tacer (1997); Picinbono (1997); Vakman (1996); Maragos et al. (1993); Boashash (1992a;1992b); Mandel (1974); Dugundji (1958)].
This paper shows the non-uniqueness of rIF obtained from different theories, defines the unique instantaneous frequency to be the cIF of the curvature radius of a signal’s trajectory on the phase plane, provides fundamental understanding of cIF, and compares cIF and rIF from different time–frequency analysis methods, including the well-known Hilbert–Huang transform (HHT) [Huang and Shen (2005)], the Teager–Kaiser algorithm (TKA) [Maragos et al. (1993)], and a conjugate-pair decomposition (CPD) method that improves a newly developed three-point frequency tracking method [Pai (2009a)].

2. Instantaneous Frequency

The decomposition of a given time function \( u(t) \) into a time-varying amplitude \( a(t) \) and a time-varying phase \( \theta(t) \) such that \( u(t) = a(t) \cos \theta(t) \) is not unique because one equation cannot be solved for two unknowns. The most popular way of finding another equation in order to have a unique decomposition is to define the conjugate part \( v(t) \) using the Hilbert transform (HT) as \( v \equiv HT(u) \). If

\[
\begin{align*}
v &\equiv HT(a \cos \theta) = a HT(\cos \theta), \\
HT(\cos \theta) &\equiv \sin \theta,
\end{align*}
\]

then one can form the complex function \( z(t) \equiv u + jv = ae^{j\theta} \) with \( j \equiv \sqrt{-1} \) and obtain the time-varying amplitude \( a \) and \( \omega \) as:

\[
\begin{align*}
a &\equiv \sqrt{u^2 + v^2}, \\
\omega &\equiv \dot{\theta} = \frac{d}{dt}(\tan^{-1}(a \sin \theta/a \cos \theta)) = \frac{d}{dt}(\tan^{-1} \frac{v}{u}) = \frac{u \dot{v} - \dot{u} v}{u^2 + v^2}. \\
\end{align*}
\]

Unfortunately, the Bedrosian theorem [Bedrosian (1963)] only guarantees the validity of Eq. (1a) for a signal with its frequency band of \( a(t) \) being lower than and non-overlapped with its frequency band of \( \cos \theta(t) \). For general nonlinear nonstationary signals, Eq. (1a) is not guaranteed. Moreover, for a general nonlinear time function \( \theta(t) \), Eq. (1b) may not be valid (i.e. the Nuttall theorem [Nuttall (1966)])

However, for signals having constant amplitudes and frequencies, Eqs. (1a)–(1c) are always exact and the obtained frequency is a circular frequency because it represents the counterclockwise angular speed along the signal’s circular path on the phase plane.

Because the decomposition of a time function \( u(t) \) into \( u(t) = a(t) \cos \theta(t) \) is not unique, different decompositions result in different rIF values. To demonstrate the non-uniqueness of rIF we consider a pure harmonic signal \( u(t) = \cos \Omega t \) having a constant frequency \( \Omega (= 2\pi) \) and rewrite it into an AM-FM signal as:

\[
\begin{align*}
u &= \cos \Omega t = ax, \\
a &\equiv \frac{1 + \varepsilon}{1 + \varepsilon \cos n\Omega t}, \\
x &\equiv \frac{1 + \varepsilon \cos n\Omega t}{1 + \epsilon} \cos \Omega t, \\
y &\equiv \sqrt{1 - x^2}, \\
\theta &\equiv \tan^{-1} \frac{yS}{x}, \\
\omega &\equiv \dot{\theta} = (x\dot{y} - \dot{x} y) s, \\
v &\equiv ays \neq a HT(x) \neq HT(u),
\end{align*}
\]
where $\varepsilon$ is a small parameter and $s \equiv \text{sign}(x\dot{y} - \dot{x}y)$ is used to ensure $\omega \geq 0$. The definition $\omega \equiv \dot{\theta}$ implies that the conjugate part is assumed to be $v(t) \equiv ays$ (see Eq. (1c)) and the angular speed of the position vector from the origin to the point $(u, v)$ on the phase plane is taken to be the instantaneous frequency. The $\omega$ from Eq. (2b) is not a circular frequency because the instantaneous trajectory is not a circular arc with a center at $(u, v) = (0, 0)$. For $\varepsilon = 0.2$ and $n = 2$, Figs. 1(c) and 1(d) show that the frequency $\omega$ is not a circular frequency because the trajectory on the phase plane is not a circular one. This decomposition violates the Bedrosian theorem because the frequency bands of $a$ and $x$ overlap, but $ax$ still represents the

Fig. 1. Two different decompositions of a regular harmonic: (a) $u(t)$, (b) amplitudes, (c) frequencies, and (d) trajectories on the phase plane.
exact \( u \). If \( 0 < u < 0.5 \), one can show that the decomposition satisfies the Bedrosian theorem, but the obtained frequency is also not a circular frequency. Only when \( n = 0, v = HT(u) \) and \( \omega(= \Omega) \) is a circular frequency because it represents the angular speed on the circular trajectory having a radius \( a = 1 \). The key point here is that the unique instantaneous frequency should be the circular frequency.

2.1. Circular instantaneous frequency of an arbitrary signal

Next we derive the circular instantaneous frequency (cIF) of an arbitrary signal. Any time function \( u(t) \) defined over \( 0 \leq t \leq T \) can be extended by repeating itself to become a periodic function, and the extended function can be expressed in terms of regular harmonics having constant amplitudes and frequencies by using the discrete Fourier transform (DFT) as [Brigham (1974)]

\[
u(t) = \sum_{i=-N/2}^{N/2} (a_i \cos \omega_i t + b_i \sin \omega_i t)
= a_0 + 2 \sum_{i=1}^{N/2} (a_i \cos \omega_i t + b_i \sin \omega_i t)
= a_0 + \sum_{i=1}^{N/2} A_i \cos \theta_i
\]

where \( \omega_i \equiv 2\pi i/T, u_k \equiv u(t_k), t_k \equiv (k-1)\Delta t, k = 1, \ldots, N \), \( N \) is the total number of samples (assumed to be even here), \( \Delta t \) is the sampling interval, \( 1/\Delta t \) is the sampling frequency, \( T (= N\Delta t) \) is the sampled period (or signal duration), \( \Delta f (= 1/T) \) is the frequency resolution, and \( 1/(2\Delta t)(= N/(2T)) \) is the Nyquist (or maximum) frequency. Note that \( a_{-i} = a_i, b_{-i} = -b_i, a_{-i} \cos \omega_{-i} t_k = a_i \cos \omega_i t_k, \) and \( b_{-i} \sin \omega_{-i} t_k = b_i \sin \omega_i t_k \) are used in Eq. (3). Moreover, \( a_i \) and \( b_i \) are constant amplitudes of the regular harmonics.

Because the lowest-frequency harmonic in Eq. (3) has a frequency \( \omega_1 = 2\pi/T \), the \( T \) needs to cover at least one period of the signal in order to identify the signal’s lowest-frequency component from the spectrum. Because \( u_{N+1} = u_1 \) is implicitly assumed in DFT, if sudden change (i.e. discontinuity) exists between \( u_N \) and \( u_1 \) due to sampling and/or transient response, the \( \omega_1 \) harmonic is not a real component. Moreover, in order to account for this discontinuity, many high-frequency harmonics exist and result in the so-called leakage problem in the frequency domain and the Gibbs’ phenomenon in the time domain, as shown in Fig. 2. The leakage problem and Gibbs’ effect can be eliminated by extending the two data ends with
smooth curves to have continuous \( u, \dot{u}, \text{and } \ddot{u} \) everywhere, including at \( t = T \). Of course, this process can only be done case by case, and it is difficult to generalize this signal processing process. However, it is theoretically possible to exactly represent an arbitrary continuous signal in terms of regular harmonics of constant frequencies and amplitudes by the Fourier transform with appropriate sampling and data extension. After Eq. (3) is obtained, because \( A_i \) and \( \dot{\theta}_i \) are constant, the exact conjugate part of \( u(t) \) can be obtained using HT as:

\[
v(t) = HT \left( a_0 + \sum_{i=1}^{N/2} A_i \cos \theta_i \right) = \sum_{i=1}^{N/2} A_i \sin \theta_i.
\] (4)

Then, it follows from Fig. 3 that the curvature, circular IF, and instantaneous center of the trajectory on the phase plane can be obtained as:

\[
\text{Curvature: } \frac{1}{a} = -\frac{d \dot{v}}{dt} \cdot \mathbf{i}_a = \frac{\ddot{u} \dot{v} - \ddot{v} \dot{u}}{V^3}, \quad V \equiv \sqrt{\dot{u}^2 + \dot{v}^2}, \quad \mathbf{i}_l = \frac{\dot{v}}{V} \mathbf{i} + \frac{\dot{u}}{V} \mathbf{j},
\]

\[
\mathbf{i}_a = \frac{\dot{v}}{V} \mathbf{i} - \frac{\dot{u}}{V} \mathbf{j}
\] (5)

\[
\text{Circular IF: } \omega = \frac{d \theta}{dt} = \frac{\ddot{u} \dot{v} - \ddot{v} \dot{u}}{V^2} \neq \frac{\dot{u} \dot{v} - \ddot{u} \ddot{v}}{u^2 + v^2}, \quad \theta \equiv \tan^{-1} \frac{-\dot{u}}{\dot{v}} \neq \tan^{-1} \frac{v}{u}
\]

\[
\text{Instantaneous Center: } \left( \tilde{u}, \tilde{v} \right) = (u - a \cos \theta, v - a \sin \theta),
\]

where \( \mathbf{i} \) is the unit vector along the \( u \)-axis, \( \mathbf{j} \) is the unit vector along the \( v \)-axis, \( \mathbf{i}_l \) is the unit vector tangent to the trajectory, and \( \mathbf{i}_a \) is the outward unit vector normal to the trajectory. Because Eq. (5) shows that the curvature and cIF have the same sign, if the radius of curvature is assumed to be always positive (i.e. \( a \equiv V^3/|\ddot{u} \dot{v} - \ddot{v} \dot{u}| \)), the cIF will be always positive (i.e. \( \omega = |\ddot{u} \dot{v} - \ddot{v} \dot{u}|/V^2 \)).

Fig. 2. The HT and IDFT of \( u(t) = \cos 2\pi t + 0.1 \cos 6\pi t \) with \( 0 < t < T(= 2.85) \) and \( N = 100 \), where the solid black lines represent the exact \( u(t) \) and \( v(t) \), the dots and the \( u \) curve represent the \( u(t) \) from IDFT, and the dots and the \( v \) curve represent the \( v(t) \) from HT.
To illustrate the difference between cIF and rIF, we consider the following signal $u(t)$ consisting of two regular harmonics having constant amplitudes $a_1$ and $a_2$:

$$u(t) = a_1 \cos \theta_1 + a_2 \cos \theta_2,$$

where $\theta_1 \equiv \omega_1 t$, $\theta_2 \equiv \omega_2 t$, and $\omega_1$ and $\omega_2$ are constant. It follows from Fig. 4 and the law of cosines that

$$a_{12} = \sqrt{a_1^2 + a_2^2 + 2a_1a_2 \cos(\theta_1 - \theta_2)}, \quad \theta_{12} = \tan^{-1} \frac{a_1 \sin \theta_1 + a_2 \sin \theta_2}{a_1 \cos \theta_1 + a_2 \cos \theta_2},$$

$$\omega_{12} \equiv \dot{\theta}_{12} = \frac{1}{2}(\omega_1 + \omega_2) + \frac{1}{2}(\omega_1 - \omega_2) \frac{a_1^2 - a_2^2}{a_{12}^2}.$$
On the other hand, it follows from Eqs. (5) and (6a) that

\[
\omega = \frac{\dddot{u} - \ddot{v}}{V^2} = \frac{\omega_1^2 a_1^2 + \omega_2^2 a_2^2 + (\omega_1 + \omega_2)\omega_1 \omega_2 a_1 a_2 \cos(\theta_1 - \theta_2)}{\omega_1^2 a_1^2 + \omega_2^2 a_2^2 + 2\omega_1 \omega_2 a_1 a_2 \cos(\theta_1 - \theta_2)}.
\]

(6d)

If \( a_1 = a_2 \) and \( \omega_2 \gg \omega_1 \approx 0 \), \( u(t) \approx \bar{u} + a_2 \cos \omega_2 t \) with \( \bar{u} \equiv a_1 \cos \theta_1 \) being a constant around \( t = t_n \). The cIF from Eq. (6d) is \( \omega \approx \omega_2 \) because the instantaneous trajectory circles around Point \( O \) in Fig. 4 with a frequency close to \( \omega_2 \). On the other hand, it follows from Eq. (6c) that \( \omega_{12} = (\omega_1 + \omega_2)/2 \approx \omega_2/2 \), which is the rIF with respect to Point \( O_1 \). Because \( \omega_1 \approx 0 \), the period \( 2\pi/\omega_1 \rightarrow \infty \) and any finite-length sampling will present the signal as \( u(t) \approx a_1 + a_2 \cos \omega_2 t \) around \( t = 0 \). Hence, \( v(t) = HT(u(t)) \approx a_2 \sin \omega_2 t \), and the actual frequency should be close to \( \omega_2 \). Obtaining \( v = HT(u) = a_1 \sin \theta_1 + a_2 \sin \theta_2 \) requires an infinite length of sampling in order to present the signal as \( u(t) = a_1 \cos \theta_1 + a_2 \cos \theta_2 \) and to have a zero global average \( a_0 \), which is almost impossible to be numerically realized. The erroneous prediction of Eq. (6c) results from the assumptions \( u(t) = a_1 \cos \theta_1 + a_2 \cos \theta_2 \) and \( v(t) = a_1 \sin \theta_1 + a_2 \sin \theta_2 \) and the use of Point \( O_1 \) as the instantaneous trajectory center. If the slow component \( a_1 \cos \omega_1 t \) is known and temporarily fixed, the modified trajectory will center at point \( O_2 \) and the frequency will be \( \omega_2 \). The main point here is that, to obtain a meaningful IF, the slow moving average \( \bar{u} \) needs to be known and removed from \( u \) or the instantaneous center \( O \) needs to be found. All the mistakes and paradoxical theories about IF in the literature are mainly caused by the co-existence of these three centers \( O, O_1, \) and \( O_2 \) and the difficulty in tracking them on the time/frequency domains.

To show a numerical example of cIF we consider

\[
u = \cos(2\pi t) + 0.5 \cos(4\pi t) + 0.5 \cos(6\pi t).
\]

(7)

Figures 5(b) and 5(d) show that the curvature radius \( a \) changes continuously, and Figs. 5(a) and 5(d) show that the instantaneous center \( \bar{u} \) changes with time. Moreover, Figs. 5(c) and 5(d) show that the cIF is positive because the curvature radius rotates along the counterclockwise direction at any time. On the other hand, the rIF with respect to the origin is not a circular frequency and it becomes negative when the position vector rotates along the clockwise direction, as shown in Fig. 5(d). The main point is that a physically meaningful and unique cIF can be defined for any signal even if it contains a nonzero moving average, an asymmetric envelope, and/or multiple independent components at any time instant. Moreover, it will be shown later in Sec. 4 that this concept of cIF enables the development of online frequency tracking methods for an arbitrary signal.

This process decomposes a signal into an instantaneous harmonic component and an instantaneous center, but the curvature radius and the moving center are difficult to be accepted as the signal’s amplitude and moving average. Hence, it cannot be used for signal decomposition. To demonstrate this point and show numerical examples about the difference between cIF and rIF we consider the following AM
Fig. 5. A signal consisting of three regular harmonics: (a) $u(t)$, (b) radius of curvature, (c) circular frequency, and (d) the trajectory and radius of curvature on the phase plane.

The signal

$$u(t) = (1 + \varepsilon \cos 2\Omega t) \cos \Omega t,$$

where $\varepsilon$ is a small parameter. One may intuitively assume that

$$u(t) = a_1 \cos \theta_1, \quad a_1 \equiv 1 + \varepsilon \cos 2\Omega t, \quad \theta_1 \equiv \Omega t, \quad \omega_1 \equiv \dot{\theta}_1 = \Omega,$$

$$v(t) = a_1 \sin \theta_1 = a_1 HT(\cos \theta_1) \neq HT(u).$$

(9a)

Defining $\omega_1 \equiv \dot{\theta}_1$ implies that the conjugate part is assumed to be $v(t) \equiv a_1 \sin \theta_1$ without considering the validity of $HT(a_1 \cos \theta_1) = a_1 HT(\cos \theta_1)$.
Without using any transform or approximation, one can rewrite Eq. (8) into an AM-FM signal as
\[
u(t) = (1 + 0.5\varepsilon) \cos \Omega t + 0.5\varepsilon \cos 3\Omega t
\]
\[= (1 + 0.5\varepsilon + 0.5\varepsilon \cos 2\Omega t) \cos \Omega t - 0.5\varepsilon \sin 2\Omega t \sin \Omega t
\]
\[= a_2 \left( 1 + 0.5\varepsilon + 0.5\varepsilon \cos 2\Omega t \right) \cos \Omega t - \frac{0.5\varepsilon \sin 2\Omega t}{a_2} \sin \Omega t \]
\[= a_2 \cos \theta_2
\]
\[a_2 \equiv \sqrt{(1 + 0.5\varepsilon + 0.5\varepsilon \cos 2\Omega t)^2 + (0.5\varepsilon \sin 2\Omega t)^2}
\]
\[
\omega \equiv \theta_2 \equiv \Omega + \Omega \tan^{-1} \frac{0.5\varepsilon \sin 2\Omega t}{1 + 0.5\varepsilon + 0.5\varepsilon \cos 2\Omega t}
\]
\[= \tan^{-1} \frac{\varepsilon t}{u}
\]
\[v(t) \equiv a_2 \sin \theta_2 = (1 + 0.5\varepsilon) \sin \Omega t + 0.5\varepsilon \sin 3\Omega t = HT(u).
\]
Again, defining \(\omega \equiv \theta_2\) implies that the conjugate part is assumed to be \(v(t) \equiv a_2 \sin \theta_2\) without considering whether \(HT(a_2 \cos \theta_2) = a_2 HT(\cos \theta_2)\) and \(HT(\cos \theta_2) = \sin \theta_2\). However, all the derivations in Eq. (9b) are exact, and they result in \(HT(a_2 \cos \theta_2) = a_2 \sin \theta_2\) even though the frequency bands of \(a_2\) and \(\cos \theta_2\) are overlapped. In other words, the Bedrosian theorem is only a sufficient condition for \(HT(a_2 \cos \theta_2) = a_2 HT(\cos \theta_2)\).

One can also use Huang’s direct quadrature method [Huang et al. (2009)] to obtain
\[
u(t) = a_3 \cos \theta_3, \quad a_3 \equiv 1 + \varepsilon, \quad \cos \theta_3 \equiv \left(1 + \varepsilon \cos 2\Omega t\right) \cos \Omega t / (1 + \varepsilon) \equiv x
\]
\[y_3 \equiv \sqrt{1 - x^2}, \quad \theta_3 \equiv \tan^{-1} \frac{y_3}{x}, \quad \omega_3 \equiv \theta_3 = (x\dot{y}_3 - \dot{x}y_3)s,
\]
\[v(t) \equiv a_3 y_3 s = a_3 \sin \theta_3 \neq HT(u),
\]
where \(s \equiv \mathrm{sign}(x\dot{y}_3 - \dot{x}y_3)\) is used to ensure \(\omega_3 \geq 0\). This method does not suffer from the edge effect shown in Fig. 2 because the conjugate part is not obtained from HT. However, \(y_3 s = \sin \theta_3 \neq HT(\cos \theta_3)\) because \(x^2 + y^2 = 1\) but \(x^2 + (HT(x))^2 \neq 1\). Hence, \(v \neq HT(u)\), and \(\omega_3\) is different from \(\omega_2\) because \(\theta_3\) is distorted and different from \(\theta_2\). One can also use the direct inverse of cosine to obtain the same IF as \(\omega \equiv d(\cos^{-1} x)/dt = -\dot{x}s/\sqrt{1 - x^2}\) where \(s \equiv \mathrm{sign}(-\dot{x})\) is used to ensure \(\omega_3 \geq 0\).

One can also use Huang’s normalized Hilbert transform [Huang et al. (2009)] to obtain:
\[
u(t) = a_4 x = a_4 \sqrt{x^2 + y^2} \cos \theta_4, \quad a_4 = 1 + \varepsilon,
\]
\[x \equiv (1 + \varepsilon \cos 2\Omega t) \cos \Omega t / (1 + \varepsilon), \quad y \equiv HT(x),
\]
\[\theta_4 \equiv \tan^{-1} \frac{y}{x}, \quad \omega_4 \equiv \theta_4 = \frac{x\dot{y} - \dot{x}y}{x^2 + y^2}, \quad v(t) \equiv a_4 y = a_4 \sqrt{x^2 + y^2} \sin \theta_4.
\]
Because of the normalization of amplitude, this method can reduce the sampling-induced discontinuity (between the two data ends) and the edge effect although HT is used. Note that $x^2 + y^2 \neq 1$ and $HT(x) = \sqrt{x^2 + y^2}\sin \theta_4 \neq \sin \theta_4$. Because the actual amplitude $a_4\sqrt{x^2 + y^2}$ is not constant, $\omega_4$ is not a cIF. For this specific case, $v = a_4HT(x) = HT(u)$ because $a_4$ is constant. If $a_4$ is not constant, $v \neq HT(u)$ and $\omega_4 \neq \omega_2$ may occur.

Of course the signal can also be decomposed into

$$u(t) = a_5\cos \theta_5 + \tilde{a}, \quad \omega_5 \equiv \dot{\theta}_5, \quad v(t) \equiv HT(u),$$

where $a_5$ is the instantaneous curvature radius, and $\omega_5$ is the unique cIF. Unfortunately, $a_5$ and $\tilde{a}$ cannot be accepted as the signal’s amplitude and moving average, as shown in the next three numerical examples.

For $\Omega = 2\pi$ and $\varepsilon = 0.1$, Figs. 6(b) and 6(c) show that these five sets of amplitudes and frequencies do not coincide, but each of Eqs. (9a)–(9e) results in the exact $u(t)$ shown in Eq. (8) and each of the IFs in Fig. 6(c) has an average frequency of 1 Hz over each period. In general, the trajectories (on the $u$–$v$ plane) from different methods do not coincide because of different ways of defining the conjugate part, as shown in Fig. 6(d). In Eq. (9c), $\omega_3$ is a circular frequency because $u^2 + v^2 = a_3^2 = (1 + \varepsilon)^2$. However, for a general nonstationary signal, $\omega_3$ is not a circular frequency because $a_3$ may change with time. Hence, in general, $\omega_i (i = 1, 2, 3, 4)$ are rIF because the corresponding trajectories are not circular ones with a center at $(u, v) = (0, 0)$. On the other hand, $\omega_5$ is the exact cIF because it represents the actual angular speed with respect to the instantaneous center at $(u, v) = (\tilde{u}, \tilde{v})$ on the phase plane, as shown in Fig. 6(d). This example clearly demonstrates that rIFs are not unique because they depend on how the signal’s conjugate part is defined. In Fig. 6(c), $\omega_3$ is different from $\omega_4$ because $HT(x) \neq y_3$ due to the fact that $x^2 + y_3^2 = 1$, but $x^2 + y^2 \neq 1$. In Fig. 6(b), if $a_4$ replaces with $a_4\sqrt{x^2 + y^2}$, the amplitude agrees with $a_2$. One may consider the use of $\tilde{a}$ to define the unique conjugate part if both $u$ and $\tilde{a}$ are measured, but a unique set of $a(t)$ and $\theta(t)$ still cannot be determined because two new unknowns $\tilde{a}(t)$ and $\dot{\theta}(t)$ are introduced in addition to the two unknowns $a(t)$ and $\theta(t)$.

Because the signal of Eq. (8) is nothing but a harmonic $(1 + 0.5\varepsilon)\cos \Omega t$ slightly distorted by $0.5\varepsilon \cos 3\Omega t$, one would expect the frequency to be close to $\Omega$ and the amplitude to be close to $1 + \varepsilon/2$. Unfortunately, the instantaneous center $\tilde{u}$, curvature radius $a_5$, and circular frequency $\omega_5$ are difficult to be accepted as the moving average, amplitude, and frequency, respectively. For example, if $\Omega = 2\pi$ and $\varepsilon = 0.4$, Figs. 7(b) and 7(c) show that the curvature radius and circular frequency can be negative. A negative frequency simply means that the curvature radius rotates along the clockwise direction on the phase plane, which is nothing wrong. Although one can assume the curvature radius is always positive and hence $\omega_5$ is also positive, the corresponding $\omega_5$ does not have the intuitively expected (but not mathematically required) average of 1 Hz for each period. Moreover, the significant variation and jumps of $\tilde{u}$ in Fig. 7(a) indicate that the use of circular frequency and
the decomposition shown in Eq. (9e) may be mathematically rigorous but it deviates too much from our expectations for weakly nonlinear signals, which should be close to the corresponding linear ones. Hence, Eq. (9e) is not intuitively meaningful and useful in signal processing for system identification of linear and nonlinear systems, as shown later in Sec. 2.2.

If \( \Omega = 2\pi \) and \( \varepsilon = 1.8 \), Figs. 8(a)–8(d) show that, when the time trace \( u(t) \) has extra local minima and maxima and hence the trajectory has extra loops on the phase plane, \( \omega_4 (= \omega_2) \) can be negative when the position vector from the origin...
rotates along the clockwise direction on the phase plane. However, the existence of extra local minima and maxima on the time trace often implies a multiple-mode response of a mechanical system of multiple degrees of freedom, and hence modal decomposition is needed before further signal processing for system identification [Huang and Shen (2005); Huang et al. (2009); Huang and Attoh-Okine (2005)]. After modal decomposition, the IF of each decomposed modal function should be positive if its trajectory does not have extra loops.
Fig. 8. Different decompositions of Eq. (8) with $\varepsilon = 1.8$: (a) $u(t)$, (b) amplitudes, (c) frequencies, and (d) trajectories and curvature radi.

Figures 6(d), 7(d), and 8(d) show that the trajectories of Eqs. (9d) and (9e) from the use of HT coincide with the exact one in Eq. (9b). Moreover, the frequency and amplitude from Eq. (9e) are far different from intuitive expectations. Hence, the use of HT to define a signal’s conjugate part is believed to be the most suitable approach, and the use of $(u, v) = (0, 0)$ as the reference point for calculating IF (even
if the calculated IF is not a circular frequency) is more convenient and useful for practical applications. However, the resulted rIF is better to be non-negative, which requires no loops on the phase plane (see Figs. 5(d) and 8(d)) and no local maxima or minima within each lowest frequency period in the time domain (see Figs. 5(a) and 8(a)). Otherwise, signal decomposition is needed. Moreover, as discussed using Fig. 4, any obvious moving average needs to be removed before a meaningful rIF can be calculated. All these requirements for signal decomposition and the calculation of meaningful rIF are exactly what HHT is proposed to do for signal processing of general nonlinear and/or nonstationary dynamical signals, as shown later in Sec. 3.

2.2. Physically meaningful referred instantaneous frequency

To show an example of physically meaningful rIF, we consider the following Duffing oscillator subjected to a harmonic excitation at a frequency \( \Omega \) close to its linear natural frequency \( \omega \) and the corresponding steady-state second-order perturbation solution [Nayfeh and Mook (1979)]:

\[
\ddot{u} + 2\varsigma \omega \dot{u} + \omega^2 u + \alpha u^3 = F \cos \Omega t
\]  

(10a)

\[
u(t) = a_1 \cos(\Omega t - \phi) + a_3 \cos(3\Omega t - 3\phi) = a(t) \cos \theta(t), \quad a_3 \equiv \frac{\alpha a_1^3}{32\omega^2} \ll a_1
\]  

(10b)

\[a(t) \equiv \sqrt{a_1^2 + a_3^2 + 2a_1a_3 \cos(2\Omega t - 2\phi)} \approx a_1 + a_3 \cos(2\Omega t - 2\phi)
\]  

(10c)

\[\theta(t) \equiv \Omega t - \phi + \tan^{-1} \frac{a_3 \sin(2\Omega t - 2\phi)}{a_1 + a_3 \cos(2\Omega t - 2\phi)} \approx \Omega t - \phi + \frac{a_3}{a_1} \sin(2\Omega t - 2\phi)
\]  

(10d)

\[\hat{\Omega} \equiv \dot{\theta} = \Omega + \frac{2\Omega a_3^2 + 2\Omega a_1 a_3 \cos(2\Omega t - 2\phi)}{a_1^2 + a_3^2 + 2a_1a_3 \cos(2\Omega t - 2\phi)} \approx \Omega + \frac{2\Omega a_3}{a_1} \cos(2\Omega t - 2\phi)
\]  

(10e)

\[
\left( \frac{3\alpha}{8\omega} \right)^2 a_1^2 - \frac{3\alpha \sigma}{4\omega} a_1^4 + (\sigma^2 + \varsigma^2 \omega^2) a_1^2 - \left( \frac{F}{2\omega} \right)^2 = 0, \quad \sigma \equiv \Omega - \omega
\]  

(10f)

\[
\phi = \tan^{-1} \frac{\varsigma \omega}{3\alpha a_1^2/(8\omega) - \sigma}
\]  

(10g)

\[\dot{\omega} = \omega + \frac{3\alpha}{8\omega} a_1^2\]

(10h)

where \( \phi \) is the phase difference between the forcing function and the response, and \( \varsigma, \alpha, F, \) and \( \phi \) are constants. The amplitude \( a_1 \) is a nonlinear function of \( F, \varsigma, \omega, \) and \( \alpha \), as shown in Eq. (10f) from perturbation analysis [Nayfeh and Mook (1979)]. Equation (10b) shows that \( u(t) \) consists of two synchronous harmonics (i.e. reaching maxima at the same time), and Eq. (10b) is similar to Eq. (9b). Because \( a_1 \gg a_3 \), \( u(t) \) would appear as one distorted harmonic (i.e. a harmonic with time-varying amplitude and frequency) with an amplitude \( a \) and a frequency \( \hat{\Omega} \) varying at a frequency \( 2\Omega \). This phenomenon can be used to determine the order (cubic
or other) of nonlinearity. Moreover, if $\alpha > 0$, $a_3/a_1 > 0$, and $\Omega$ and $a$ are at their maxima when $u(t)$ is at its maxima or minima. If $\alpha < 0$, $a_3/a_1 < 0$, and $\Omega$ and $a$ are at their minima when $u(t)$ is at its maxima or minima. This phenomenon can be used to determine the type (hardening or softening) of nonlinearity, and the magnitude of nonlinearity can be estimated using Eq. (10b) and the time-domain data $a(t)$ as $\alpha = 32\Omega^2 a_3/a_1$. Moreover, one can use Eq. (10g) to estimate $\varsigma$, use Eq. (10h) to estimate the undamped amplitude-dependent natural frequency $\hat{\omega}$, and use Eq. (10f) to obtain $F$ and then estimate the mass as $m = F_0/F$, where $F_0$ is the known excitation amplitude. The key point here is that one steady-state response of a nonlinear system to a harmonic excitation can be used to identify all system parameters (if the time-varying frequency and amplitude can be extracted), which is impossible for linear cases.

A high-order linear dynamical system can be decomposed into multiple second-order linear systems, and linear free vibration of a second-order linear system is described by a regular harmonic function having a constant frequency. Hence, harmonic time functions are often chosen as basis functions for decomposition of time-domain signals, and a structure’s basis functions in the spatial domain are the so-called mode shapes. However, a complex system’s basis responses may not be harmonic but are physically meaningful basis functions for analyzing the system’s responses to other compound/complicated excitations. Equations (10a)–(10h) show that the steady-state response of a nonlinear system to a harmonic excitation is an amplitude- and frequency-modulated harmonic, and it is physically more meaningful to treat this distorted harmonic as one independent basis function, instead of separating it into two or more regular harmonics or decomposing it into the form of Eq. (9e). Then, the perturbation solution can be directly used with the time-varying amplitude and non-circular rIF for accurate nonlinear system identification. This concept is similar to the treatment of elastic coupling of structural vibration modes. For example, because a symmetrically laminated composite beam possesses elastic bending–torsion coupling, its bending modes are always accompanied with torsional deformation. Hence, it is meaningless to decompose such a bending–torsion mode into a bending mode and a torsional mode.

Wavelet transform can be used to decompose a linear system’s response signal into basis functions that represent independent characteristic responses of the system. Unfortunately, the frequencies of basis functions of nonlinear systems vary with time, but wavelet transform uses predetermined basis functions (i.e. wavelets) and hence cannot adapt and track such basis functions.

3. Extraction of Meaningful Referred Instantaneous Frequency

Hilbert–Huang transform (HHT) uses the empirical mode decomposition (EMD) to decompose a nonlinear nonstationary signal into a few intrinsic mode functions (IMFs), and then uses Hilbert transform (HT) to calculate each IMF’s time-varying
rIA and rIF. EMD is a data-driven signal decomposition technique that sequentially extracts zero-mean IMFs from a signal, starting from high- to low-frequency components, and it is a dyadic filter equivalent to an adaptive wavelet. HHT is essentially different from Fourier and wavelet transforms because it does not use pre-determined basis functions and the convolution between the basis functions and the signal itself to extract components. The first step of HHT is to use EMD to sequentially decompose a signal \( u(t) \) into \( n \) IMFs \( c_i(t) \) and a residual \( r_n \) as [Huang and Shen (2005)]

\[
u(t) = \sum_{i=1}^{n} c_i(t) + r_n(t), \quad (11)\]

where \( c_1 \) has the shortest characteristic time scale 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 the local maxima using a natural cubic spline, compute the lower envelope by connecting all the local minima using another natural cubic spline, subtract the mean of the upper and lower envelopes, \( m_{11} \), 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. \( m_{1K} \approx 0 \)), and then define \( c_1 \) as:

\[
c_1 \equiv u - m_{11} \cdots - m_{1K}. \quad (12)\]

This sifting process eliminates low-frequency riding waves (e.g. \( \bar{u}(t) \) in Fig. 4), makes the wave profile symmetric, and separates the highest-frequency IMF from the current residuary signal. During the sifting process for each IMF a deviation \( D_v \) is computed from the two consecutive sifting results as:

\[
D_v \equiv \sqrt{\sum_{i=1}^{N} (c_{1k}(t_i) - c_{1k-1}(t_i))^2 / \sum_{i=1}^{N} c_{1k-1}^2(t_i)}, \quad (13)\]

where, for example, \( c_{1k} \equiv u - m_{11} \cdots - m_{1k}, \) \( t_i = (i-1)\Delta t, \) and \( T(=N\Delta t) \) is the sampled period. A systematic method to end the iteration is to limit \( D_v \) to be a small number and/or to limit the maximum number of iterations. After \( c_1 \) is obtained, define the residual \( r_1(\equiv u - c_1) \), treat \( r_1 \) as the new data, and repeat the steps shown in Eq. (12) to obtain other \( c_i (i=2, \ldots, n) \) as:

\[
c_n = r_{n-1} - m_{n1} \cdots - m_{nK}, \quad r_{n-1} \equiv u(t) - c_1 \cdots - c_{n-1}. \quad (14)\]

The whole sifting process can be stopped when the residual \( r_n (= r_{n-1} - c_n) \) 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 and has at most one local extremum.

The second step of HHT is to perform Hilbert transform and compute the time-varying rIF \( \omega_i \) and amplitude \( A_i \) of each \( c_i \). After all \( c_i(t) \) are extracted, one can perform Hilbert transform (see Eqs. (3) and (4)) to obtain \( d_i(t) \) from each \( c_i(t) \) and

\[
\text{Circular Instantaneous Frequency} \quad 55
then define \( z_i(t) \) and use Eq. (11) with \( r_n \) being neglected to obtain

\[
z_i(t) \equiv c_i(t) + j d_i(t) = A_i e^{j \theta_i}, \quad A_i \equiv \sqrt{c_i^2 + d_i^2},
\]

\[
\theta_i \equiv \tan^{-1} \frac{d_i}{c_i}, \quad \omega_i \equiv \dot{\theta_i}
\]

\[
u(t) = \text{Re} \left( \sum_{i=1}^{n} [c_i(t) + j d_i(t)] \right) = \text{Re} \left( \sum_{i=1}^{n} A_i(t)e^{j \theta_i(t)} \right).
\] (15)

The time-varying \( A_i(t) \) and \( \theta_i(t) \) in Eq. (15) indicate that HHT is a time–frequency analysis technique, and it often decomposes a complex nonlinear nonstationary signal into just a small number of IMFs because distorted harmonics are allowed. Because the moving average of each \( c_i(t) \) is removed by the EMD process, the instantaneous frequency \( \omega_i \) from Eq. (15) is a physically meaningful rIF.

Because of the capability of accurate extraction of time-varying frequencies and amplitudes, HHT with perturbation analysis (see Sec. 2.2) is an excellent combination that can be used for accurate noise filtering, capture of transient events, parametric identification, non-parametric identification, extraction of nonlinear phenomena, and identification of nonlinear continuous systems [Pai and Palazzotto (2008a;2008b); Pai (2009b)]. Although an IMF may be different from a modal vibration function due to signal processing errors and/or modal coupling, EMD is probably the best approach at present time for signal decomposition of nonlinear nonstationary signals.

For the signal shown in Fig. 1(a), one can easily point out that the average frequency is 1.0 Hz and the average amplitude is 1.0. But, if only four or even just three consecutive data points are given (see Fig. 1(a)), how to estimate the instantaneous frequency and amplitude is challenging, which is the so-called frequency tracking problem. Because HHT requires a certain length of data with certain number of maxima and minima and the use of HT, it cannot be used for frequency tracking because a set of three or four data points is too short to have maxima and minima and to perform HT.

The use of a signal’s trajectory for the calculation of cIF requires the use of HT to calculate the signal’s conjugate part, but HT is valid only if the data length is long enough to cover a few periods of the signal. Hence, this approach cannot be used for online frequency tracking, where the number of data points to be processed is limited by the processing speed of hardware and software, possible existence of fast time-varying nonlinear and/or nonstationary effects, and availability of data points. Fortunately, the concept of circular instantaneous frequency enables the development of numerical methods that use only a few data points for frequency tracking, as shown next.

4. Online Tracking of Circular Instantaneous Frequency

To understand the problems and challenges in frequency tracking, we compare the Teager–Kaiser Algorithm [Maragos et al. (1993)] and a conjugate-pair
decomposition method that extends the capabilities of a newly developed 3-point frequency tracking method [Pai (2009a)].

4.1. Teager–Kaiser algorithm

To estimate the circular frequency of a signal \( u(t) \) at an arbitrary time one can assume it to be locally harmonic and hence

\[
\begin{align*}
  u &\equiv a \cos \omega t , \quad \dot{u} = -\omega a \sin \omega t , \\
  \ddot{u} &\equiv -\omega^2 a \cos \omega t , \quad \dddot{u} = \omega^3 a \sin \omega t , \quad u^{iv} = \omega^4 a \cos \omega t ,
\end{align*}
\]

where \( u^{iv} \equiv d^4 u/dt^4 \). Three energy-type functions \( \psi(u), \psi(\ddot{u}), \) and \( \psi(\dot{u}) \) can be derived from Eq. (16) as:

\[
\begin{align*}
  \psi(u) &\equiv \dddot{u}^2 - u \dddot{u} = \omega^2 a^2 , \\
  \psi(\ddot{u}) &\equiv \dddot{u}^2 - \ddot{u} \dddot{u} = \omega^4 a^2 , \\
  \psi(\dot{u}) &\equiv \dddot{u}^2 - \dot{u} \dddot{u} = \omega^6 a^2 .
\end{align*}
\]

Then, the signal’s instantaneous frequency and amplitude can be calculated from these positive functions as:

\[
\begin{align}
  \omega &= \sqrt{\psi(u)/\psi(\ddot{u})} , & a &= \psi(u)/\sqrt{\psi(\dddot{u})} . \quad (18a) \\
  \omega &= \sqrt{\psi(\dot{u})/\psi(u)} , & a &= [\psi(\ddot{u})]^{3/2}/\psi(\dot{u}) . \quad (18b)
\end{align}
\]

If the analytical form of \( u(t) \) is known, one can use Eqs. (17) and (18b) to estimate the values of \( \omega \) and \( a \). Then, the instantaneous center \( \ddot{u}(t) \) and the instantaneous harmonic part \( \dddot{u}(t) \) can be estimated using

\[
\ddot{u} \equiv u - \ddot{u} , \quad \dddot{u} \equiv a \cos \left( \tan^{-1} \frac{-\omega \dddot{u}}{-\ddot{u}} \right) . \quad (18c)
\]

It follows from Eq. (18c) and \( v \equiv a \sin \omega t \) that

\[
\omega = \frac{d}{dt} \tan^{-1} \left( \frac{v}{u} \right) = \frac{d}{dt} \tan^{-1} \left( \frac{\omega^2 u}{\omega^4 u} \right) = \frac{d}{dt} \tan^{-1} \left( \frac{-\omega \dddot{u}}{-\ddot{u}} \right) = \frac{\omega(\dddot{u}^2 - \ddot{u} \dddot{u})}{\omega^4 a^2} = \frac{\psi(\dot{u})}{\omega^2 a^2} . \quad (19)
\]

Equations (17) and (19) reveal that Eq. (18b) is based on the concept of circle fitting on the phase plane because substituting \( v = a \sin \omega t \) and \( V = a \omega \) into Eq. (5) yields \( \omega = (\dddot{u}^2 - \ddot{u} \dddot{u})/\omega^4 a^2 \). Because \( a \) and \( \omega \) are assumed to be constant here, the obtained \( \omega \) can be different from the true circular frequency from Eq. (5). Moreover, because both \( a \) and \( \omega \) are assumed to be non-negative in Eqs. (18a) and (18b), which correspond to using \( a \equiv V^3/|\dddot{u} - \ddot{u} \dddot{u}| \) and \( \omega = |\dddot{u} - \ddot{u} \dddot{u}|/V^2 \) in Eq. (5).

Unfortunately, only one of \( u_n, \dot{u}_n, \ddot{u}_n, \) and \( u_n^{iv} \) is often measured in experiments, where \( u_n \equiv u(t_n) \) and \( t_n \equiv (n-1)\Delta t \). Consequently, numerical finite difference and/or integration needs to be used to compute the other four quantities at each time instant. If \( u_n (n = 1, \ldots, N) \) are measured, \( \dot{u} \) can be approximated using the two-sample backward difference \((u_n - u_{n-1})/\Delta t \) or the forward difference
(u_{n+1} - u_n)/\Delta t$, and $\bar{u}$ can be approximated using the three-sample central difference $(u_{n+1} - 2u_n + u_{n-1})/\Delta t^2$. Then, the finite difference forms of Eq. (17) are derived to be

$$
(\Delta t)^2 \psi(u_n) = (u_n - u_{n-1})(u_{n+1} - u_n) - u_n(u_{n+1} - 2u_n + u_{n-1})
= u_n^2 - u_n u_{n+1}
= a^2 \sin^2(\omega \Delta t),
$$

(20a)

$$
(\Delta t)^4 \psi(\bar{u}_n) = u_n^2 - 2u_n + u_{n-1} = (u_n - u_{n-1})^2 - (u_n - 1 - u_n)(u_{n+1} - u_n)
= 2a^2 \sin^2(\omega \Delta t)[1 - \cos(\omega \Delta t)],
$$

(20b)

$$
(\Delta t)^6 \psi(\bar{u}_n) = u_n^2 - 2u_n + u_{n-1} = (u_n + 1 - 2u_n + u_{n-1})^2
- (u_n - 2u_n + u_{n-1})(u_{n+1} + 2u_n + u_{n-1})
= 4a^2 \sin^2(\omega \Delta t)[1 - \cos(\omega \Delta t)^2].
$$

(20c)

Equation (20b) uses the backward difference. It follows from Eqs. (18a) and (20a, b) that

$$
\frac{(\Delta t)^4 \psi(u_n)}{2(\Delta t)^2 \psi(u_n)} = 1 - \cos(\omega \Delta t),
$$

$$
\omega = \frac{1}{\Delta t} \cos^{-1} \left( 1 - \frac{(\Delta t)^4 \psi(u_n)}{2(\Delta t)^2 \psi(u_n)} \right),
$$

(21a)

$$
a = \sqrt{\frac{(\Delta t)^2 \psi(u_n)}{1 - \cos^2(\omega \Delta t)}}.
$$

Similarly, it follows from Eqs. (18b) and (20b, c) that

$$
\omega = \frac{1}{\Delta t} \cos^{-1} \left( 1 - \frac{(\Delta t)^6 \psi(\bar{u}_n)}{2(\Delta t)^4 \psi(\bar{u}_n)} \right),
$$

(21b)

$$
a = \sqrt{\frac{(\Delta t)^4 \psi(\bar{u}_n)/2}{1 - \cos(\omega \Delta t) + \cos^2(\omega \Delta t)}}.
$$

Equation (21a) represents the original Teager–Kaiser algorithm (TKA) [Maragos et al. (1993)], but it does not work for signals with moving averages because the existence of any moving average invalidates the first identity in Eq. (17) [Pai (2009a)]. Moreover, because of the use of finite difference, results from TKA are inaccurate when the processed signal has high-frequency intra-wave modulation. Because $u_{n-2}, u_{n-1}, u_n$, and $u_{n+1}$ are used to estimate the frequency and amplitude at $t_n$, the original TKA is a 4-point method. Equation (21b) represents an improved version of TKA, and it works well for signals with moving averages. Because $u_{n-2}, u_{n-1}, u_n, u_{n+1}$, and $u_{n+2}$ are used to estimate the frequency and amplitude at $t_n$, this improved TKA is a 5-point method for frequency tracking.
4.2. Conjugate-pair decomposition

Next we present a 3-point method for tracking an arbitrary signal’s circular frequency. To enable frequency tracking of a signal with a nonzero instantaneous center \( C_0 \), we assume the signal to have the following form

\[
u(t) = C_0 + a \cos(\omega t + \phi) = C_0 + C \cos(\bar{\omega} t) - D \sin(\bar{\omega} t)
\]

with \( C_0, a, \omega, \) and \( \phi \) being unknown constants, \( \bar{t}(\equiv t - t_n) \) a shifted time, \( t_n \) the current time instant, and

\[
C = a \cos(\omega t_n + \phi), \quad D = a \sin(\omega t_n + \phi).
\]

To initiate this frequency tracking process starting from \( t = t_3 \), one can use Eq. (21b) of TKA to provide the initial guess of \( \omega \) for Eq. (22a) and hence there are only three unknowns \( C_0, C, a \) in Eq. (22a). The three unknowns for the data point at \( \bar{t} = 0 \) (i.e. \( t = t_n \)) can be determined by minimizing the squared error

\[
S_{\text{error}} \equiv \sum_{i=-\left(\frac{m-1}{2}\right)}^{(m-1)/2} (U_{n+i} - u_{n+i})^2,
\]

where \( m (\geq 3) \) is the total number of processed data points and is assumed to be an odd number, \( u_{n+i} \) denotes the function form shown in Eq. (22a) with \( \bar{t} = i \Delta t \), and \( U_{n+i} \) denotes the actual \( u(t) \) at \( t = t_n + i \Delta t \). Because there are only three unknowns, only three local data points (i.e. \( m = 3 \)) need to be processed in Eq. (23). After \( C \) and \( D \) are determined, the instantaneous amplitude \( a \), phase angle \( \theta \), and frequency \( \omega \) can be estimated as

\[
a = \sqrt{C^2 + D^2}, \quad \theta \equiv \omega t_n + \phi = \tan^{-1}(D/C),
\]

\[
\omega = \dot{\theta} \approx \frac{\theta(t_n) - \theta(t_n - p\Delta t)}{p\Delta t}.
\]

The backward finite difference is assumed for averaging the frequency over \( p\Delta t \) in order to reduce the influence of noise in frequency tracking. One can see from Eqs. (22b) and (24a) that \( D = HT(C) \) and \( C^2 + D^2 = a^2 \). Hence, this 3-point frequency tracking method is also based on the concept of circle fitting of data with respect to the instantaneous center on the phase plane to determine \( C_0, a, \) and \( \theta \). Noise filtering is possible if more than 3 points are used in Eq. (23). However, although more points can be used to further reduce the influence of measurement noise and/or signal processing errors, the estimated frequency is more a locally averaged value. For any point other than the starting point at \( t = t_3 \), the previous point’s frequency can be used as an initial guess of its frequency without using the TKA. Because the initial guess of \( \omega \) is updated using Eq. (24b), this method is able to adapt and capture the actual variation of frequency [Pai (2009a)]. However, because \( a \) and \( \omega \) are also assumed to be constant here, the obtained \( \omega \) is similar to that from Eq. (21b) and it may deviate from the true circular frequency from Eq. (5). Although both TKA and CPD are based on circle fitting, TKA uses finite difference and CPD uses curve fitting to implement the circle fitting process.
4.3. Numerical examples

To compare and understand the rIF and cIF obtained from HHT, TKA, and CPD, we consider the following signal consisting of two regular harmonics:

\[ u(t) = \cos 0.25\pi t + 0.8 \cos 2\pi t. \]  

(25)

This signal has a zero global average (i.e. \( a_0 = 0 \) in Eq. (3)) over \( 0 \leq t \leq 8 (= T) \), but its \( \tilde{u}(t) \) and \( \tilde{u}(t) (= \cos 0.25\pi t) \) are nonzero. With the use of \( \Delta t = 1/16 \), Figs. 9(b)

![Fig. 9. Frequency tracking of Eq. (25) using TKA: (a) \( u \) and \( \tilde{u} \), (b) amplitude, (c) frequency, and (d) trajectory and centers.](image)
and 9(c) show that the curvature radius and cIF from TKA (Eq. (21b)) slightly deviate from the analytical ones (gray lines) from Eq. (18b) because finite difference is used. If $\Delta t \leq 0.01$ is used, they coincide. However, the analytical curvature radius and cIF are also different from the exact ones from Eq. (5) because the curvature radius and circular frequency are assumed to be constant in TKA. Note that the exact curvature radius and cIF modulate uniformly at $7/8 \text{Hz}$, but those from TKA do not modulate uniformly.

Figures 10(b) and 10(c) show that the curvature radius and cIF from CPD (Eqs. (24a,b)) also deviate from the exact ones from Eq. (5) because the curvature radius and cIF are assumed to be constant for each time instant in CPD. When $\Delta t = 1/8$ is used, the curvature radius and cIF from CPD (gray lines) deviate more from the exact ones because three data points ($2\Delta t$) cover one-quarter of the period and the values are averaged over it. The curvature radius and cIF from CPD are different from those from TKA because CPD uses curve fitting and TKA uses finite difference to implement the circle fitting process.

The curvature radius and cIF shown in Figs. 9 and 10 are the amplitude and frequency of $0.8 \cos 2\pi t$ with modification caused by the derivatives of $\bar{u}(t)$ ($= \cos 0.25\pi t$) because they are based on circle fitting on the phase plane with respect to the instantaneous center $O$ (instead of $O_2$) in Fig. 4. The projection of the trajectory of point $O$ on the $u$ axis is $\tilde{u}(t)$, and the projection of the trajectory

![Fig. 10. Frequency tracking of Eq. (25) using CPD: (a) $u$ and $\tilde{u}$, (b) amplitude, and (c) frequency.](image)
of point $O_2$ on the $u$ axis is $\bar{u}(t)$. Although Figs. 9 and 10 show that the amplitude and frequency tracked by TKA are smoother than those tracked by CPD, the accuracy of TKA is easily destroyed by the existence of any noise because of the use of finite difference, but the accuracy of CPD is not sensitive to noise because of the use of curve fitting [Pai (2009a)].

Figures 9(a), 9(d), and 10(a) show that the instantaneous centers $\tilde{u}$ from TKA and CPD are closer to the moving average $\bar{u}(= \cos 0.25\pi t)$ of the upper and lower envelopes of $u$ than the exact one. However, the $\tilde{u}$ from TKA and CPD is still different from the moving average $\bar{u}$ and its frequency is not less but even higher than that of $u$. Hence, both TKA and CPD cannot be used for signal decomposition.

On the other hand, HHT analysis well separates the $0.8 \cos 2\pi t$ as the first IMF and the $\cos 0.25\pi t$ as the residual (i.e. the moving average or the riding wave).

5. Concluding Remarks

This work defines the unique instantaneous frequency of an arbitrary time signal to be the circular instantaneous frequency (cIF) of the curvature radius of the signal’s trajectory on the phase plane consisting of the signal and its conjugate part from Hilbert transform (HT). A negative cIF may happen, but it just means the curvature radius undergoes clockwise rotating, which is mathematically and physically correct.

Because a general response of a dynamical system of multiple degrees of freedom contains multiple modal vibrations, its cIF varies dramatically and is not very useful for system identification and other applications. If the signal is decomposed into modal vibration components without any moving average, each component has no local extrema within each fundamental period of the time trace and no local loops on the phase plane, each component’s referred instantaneous frequency (rIF) with respect to the origin on the phase plane is non-circular but always non-negative, and the time-varying rIF and referred instantaneous amplitude (rIA) are convenient for combining the use of perturbation analysis for system identification. In Hilbert–Huang transform, the empirical mode decomposition (EMD) decomposes a general nonlinear nonstationary signal into zero-mean intrinsic mode functions (IMFs), and HT enables accurate calculation of rIF of each IMF for system identification.

Although the concept of circular frequency is useless for signal decomposition, it enables the development of online frequency tracking methods for arbitrary signals without using HT. A 5-point frequency tracking method is developed to eliminate the incapability of the original 4-point Teager–Kaiser algorithm (TKA) for frequency tracking of signals with moving averages. Moreover, a 3-point conjugate-pair decomposition (CPD) method is derived based on circle-fitting using a pair of conjugate harmonic functions. It is shown that both CPD and TKA are based on the concept of circle fitting, but TKA uses finite difference and CPD uses curve fitting in numerical implementation. However, because finite difference is used in TKA, its accuracy is easily destroyed by noise. On the other hand, because CPD
Circular Instantaneous Frequency

is based on curve fitting, noise filtering is an implicit capability and its accuracy increases with the number of processed data points. The rIF from HHT and the cIF from CPD and TKA are different by definition. Moreover, because the instantaneous frequency and amplitude are assumed to be constant in CPD and TKA, the cIF from CPD and TKA also deviates from the exact cIF.

References


