CHAPTER 4

HHT SIFTING AND FILTERING

Reginald N. Meeson, Jr.

Time-frequency analysis is the process of determining what frequencies are present in a signal, how strong they are, and how they change over time. Understanding how the frequencies in a signal change with time can explain much about the physical processes that generate or influence the signal. Better resolution of the details of frequency changes provides better insight into these underlying physical processes.

The Hilbert–Huang transform (HHT) offers higher frequency resolution and more accurate timing of transient and non-stationary signal events than conventional integral transform techniques. The HHT separates complex signals into simpler component signals, each of which has a single, well-defined, time-varying frequency. Real-time HHT algorithms enable this enhanced signal analysis capability to be used in process monitoring and control applications.

"Sifting" is the central signal separation process of the HHT algorithm. This chapter compares the component signal separations of Huang's sifting process with those produced by filtering techniques. Although intuition seems to suggest that filtering, with appropriate real-time adjustments to parameters, could be substituted for Huang's sifting process, our results did not support this suggestion. Five case studies present HHT and filtering results for stationary amplitude- and frequency-modulated signals, as well as signals with more dynamic transient behavior. These examples show that, in general, HHT sifting and filtering separate signal components quite differently. Our experiments with example signals led to the discovery of aliasing in the HHT sifting algorithm.

4.1. Introduction

One way to describe a timed series of measurements, referred to as a "signal," is in terms of the frequencies in its variations. The process of determining what frequencies are present, how strong they are, and how they change over time is called "time-frequency analysis." Conventional time-frequency analysis techniques use integral calculus transforms to map time-based signals into frequency-based or joint time- and frequency-based representations. Examples of these techniques include Fourier transforms, windowed Fourier or Gabor transforms, wavelet transforms, and joint time-frequency distributions [see Cohen (1995) for a thorough introduction to these techniques].

The Hilbert–Huang transform (HHT) is a new time-frequency analysis technique that offers higher frequency resolution and more accurate timing of transient and non-stationary signal events than conventional Fourier and wavelet transform techniques. This approach was introduced by Huang (1998). Conventional techniques assume signals are stationary, at least within the time window of observation. Fourier analysis assumes further that the signal is harmonic and repeats itself with a period exactly matching the width of the sampling window. These analysis techniques are employed widely even though their (theoretically necessary) enabling conditions rarely hold for signals of interest.

In addition, integral transform techniques suffer from an uncertainty problem similar, mathematically, to Heisenberg's uncertainty principle in physics. This uncertainty limits their ability to accurately measure timing and frequency at the same time. That is, after a point, higher-resolution frequency measurements cannot be achieved without sacrificing timing accuracy, and vice versa. The HHT is able to resolve frequencies accurately and time them precisely without this limiting uncertainty.

The original HHT algorithm was formulated as a "batch" computation, in which a complete dataset is collected and then processed as a whole. An incremental algorithm that transforms evolving input data streams into streams of HHT results has also been developed [see Meeson (2002)]. Modern microprocessors and signal processing chips offer sufficient performance for this incremental algorithm to be used in many real-time applications. For this study the incremental algorithm served as a bridge connecting the original HHT algorithm to the incremental filtering techniques.

"Sifting" is the central signal separation process of the HHT algorithm. In the seminal work on the HHT, Huang (1998) described sifting informally as analogous to an adaptive filtering process, but then developed a different algorithmic procedure to separate signal components. This development led us to conjecture that filters, with parameters appropriately adjusted in real time, could mimic the HHT sifting process. It seemed natural to analyze the results from filtering and to compare them with the HHT results. Huang's original HHT sifting algorithm was the starting point for this comparison. The results from the original and incremental HHT algorithms are virtually identical for the example signals used in this analysis.

"Filtering," for this discussion, means conventional finite impulse response (FIR) digital filtering where filter coefficients and cutoff frequencies can be adjusted on a sample-by-sample basis. Our experiments with these signal-analysis techniques revealed new insights into the mathematical properties of the HHT signal separation process and may help refine HHT-processing techniques.

In section 4.2, we describe the objectives of the HHT signal-separation process and the desired attributes of separated components. Huang's original empirical mode decomposition algorithm, which later became known as the HHT, is described in section 4.3. Section 4.4 describes the incremental HHT algorithm and analyzes a special case where an analogy to conventional digital filtering techniques can be used. In section 4.5, we describe the shift from special-case static filtering to a general method using dynamically adjustable filters. HHT and filtering results for five example signals are compared in section 4.6. Section 4.7 concludes with a summary and some directions for future research.

4.2. Objectives of HHT sifting

The HHT sifting process separates a signal into a series of amplitude- and frequency-modulated component signals in the form

$$s(t) = \sum_i a_i(t) \cos[\varphi_i(t)] ,$$

where $a_i(t)$ represents the amplitude modulation and $\varphi_i(t)$ denotes the phase functions that represent the frequency modulation characteristics of each component.

Numerous possible solutions are available for this separation scheme. One familiar solution is the Fourier series, which is made up of constant amplitude and constant frequency (linear phase) functions [see, for example, Oppenheim and Shafer (1989) for a discussion of the Fourier series]. The solution that the HHT seeks is quite different. Rather than trying to represent a signal in terms of predetermined basis functions, the HHT tracks and adapts dynamically to transient, non-stationary, and nonlinear changes in component frequencies and amplitudes as the signal evolves over time.

Windowed Fourier and wavelet-signal-analysis techniques are also able to track slowly changing signal behavior; but, as described above, they suffer from an uncertainty problem that can limit the accuracy of the frequency (scale for wavelets) and timing information they yield. The product of the frequency (scale) variance and the timing variance for the results from these techniques has a positive lower bound. Consequently, once this limit is reached, increasing the accuracy of frequency measurements can be achieved only by sacrificing timing accuracy, and vice versa [see Cohen (1995) for a discussion of time and frequency uncertainty].

Many signals of interest contain short-duration transients that are difficult to analyze because of this uncertainty limitation. With conventional analysis techniques, it is not possible to accurately time when specific frequencies were present. Transient events can be timed accurately, but accurate frequency information cannot be resolved within that narrow time window.

HHT signal separations are not subject to this limitation and provide both accurate frequency and accurate timing simultaneously. HHT has this unique advantage over conventional time-frequency analysis techniques. HHT analysis of earthquake data, as described by Huang (2001), for example, shows a very different distribution of frequencies over time than conventional Fourier analysis. This difference may prove tremendously important in analyzing the strength of buildings, bridges, and other structures.

4.2.1. *Restrictions on amplitude and phase functions*

In order to extract the desired amplitude and frequency information, without conflicting interpretations or paradoxical results, restrictions must be imposed on the amplitude and phase functions, $a_i(t)$ and $\varphi_i(t)$. The primary requirement for HHT components is that they be sufficiently well behaved to allow extraction of well-defined amplitude and phase functions. Such functions are called "monocomponent" functions, and we distinguish them from "multicomponent" functions, from which amplitude and phase cannot be cleanly extracted. Although there seems to be no generally accepted mathematical definition of "monocomponentness," there is little debate over one primary criteria, which is that at any time a monocomponent signal must have a single, well-defined, positive instantaneous frequency represented by the derivative of its phase function.

The first approach suggested for finding the necessary conditions for a separated component's "monocomponentness" was to look at the component's analytic signal, which is given by

$$\mathcal{A}[c(t)] = c(t) + i\mathcal{H}[c(t)]\,,$$

where $c(t) = a(t)\cos[\varphi(t)]$, $\mathcal{H}(\cdot)$ is the Hilbert transform, and $i = \sqrt{-1}$ [see Cohen (1995) for a discussion of analytic signals and the Hilbert transform]. The analytic signal is a complex function whose Fourier transform is twice that of $c(t)$ over the positive frequencies and zero over negative frequencies. The spectrum of this signal, therefore, contains only positive frequencies. This fact does not guarantee, however, that the signal's instantaneous frequency (the derivative of its phase) will always be positive. Cohen (1995) shows examples of analytic signals that have paradoxical instantaneous frequency characteristics, including some with negative instantaneous frequencies. The analytic signal, therefore, by itself, does not appear to provide sufficient criteria for separating monocomponent signals.

A second approach suggested for finding monocomponent conditions was to consider the function's quadrature model:

$$\mathcal{Q}[c(t)] = a(t)e^{i\varphi(t)} = a(t)\{\cos[\varphi(t)] + i\sin[\varphi(t)]\}\,.$$

By using the additional knowledge about the Hilbert transform that $\mathcal{H}\{\cos[\varphi(t)]\} = \sin[\varphi(t)]$, the quadrature model can be compared with the analytic signal. The two are the same when the amplitude function can be factored out of the signal's Hilbert transform; i.e., when

$$\mathcal{H}\{a(t)\cos[\varphi(t)]\} = a(t)\mathcal{H}\{\cos[\varphi(t)]\} = a(t)\sin[\varphi(t)]\,.$$

The conditions under which this relationship holds were established by Bedrosian (1963) and elaborated by Nuttall (1966). The conditions are, for some positive frequency ω_0:

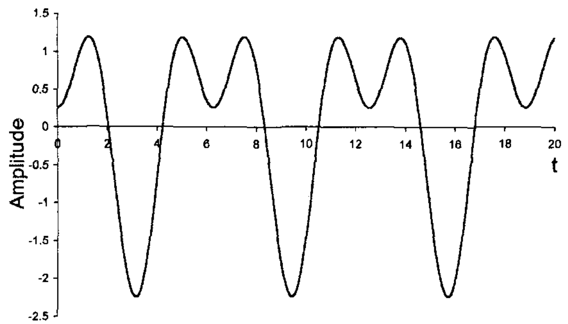

Figure 4.1: Example of a multicomponent signal.

a. The spectrum of the amplitude function is restricted to frequencies below ω_0, and
b. The spectrum of the cosine term is restricted to frequencies above ω_0.

An example of a function that does not satisfy these conditions is

$$s(t) = 1.25\cos(t) - \cos(2t)\,.$$

The analytic signal of this function is similar to one of Cohen's problematic signals, $\mathcal{A}[s(t)] = 1.25e^{it} - e^{2it}$, which cannot be expressed in the form $a(t)e^{i\varphi(t)}$ without either $a(t)$ oscillating rapidly or $\varphi'(t)$ turning negative periodically. As can be seen in the graph shown in Fig. 4.1, the real signal $s(t)$ has local minima with positive values. Such a signal cannot be expressed in the form $a(t)\cos[\varphi(t)]$ with a slowly varying amplitude and an increasing phase function. If we assume a slowly varying amplitude, to satisfy Bedrosian's first condition, then the $\cos[\varphi(t)]$ term would have to turn and go back up again without going negative. The phase function, therefore, would have to decrease for a time, resulting in a negative instantaneous frequency. This result would violate Bedrosian's spectral separation conditions, since the amplitude function would have to have a negative upper-frequency limit. If we stipulate an increasing phase function, then the amplitude must peak near $t = (2n+1)\pi$ and dip to a minimum near $t = 2n\pi$, giving an average frequency of $\omega = 1$, the same as the average change in phase. Either way, Bedrosian's condition is not satisfied.

Bedrosian's conditions are too restrictive for our needs, however. Purely frequency-modulated signals with constant amplitude can have spectra that extend down to zero frequency. Any amplitude modulation imposed on such a "carrier" signal would violate Bedrosian's conditions—even though the signal would make a perfectly good HHT component. The case studies below show that solutions must allow phase functions that exhibit this sort of frequency-modulated behavior.

Teager's energy operator, Ψ, was suggested by Maragos (1993) as a possible non-linear approach for restricting amplitude and phase functions for combined amplitude-modulated (AM) and frequency-modulated (FM) signals

$$\Psi(s(t),t) = [s'(t)]^2 - s(t)s''(t)\,.$$

For component signals of the form $a(t)\cos[\varphi(t)]$, Ψ can be expanded as

$$\begin{aligned}\Psi\{a(t)\cos[\varphi(t)],t\} = {} & [a(t)\varphi'(t)]^2 + \tfrac{1}{2}a^2(t)\sin[2\varphi(t)]\phi''(t) \\ & + \cos^2[\varphi(t)]\Psi[a(t),t]\,.\end{aligned}$$

If a signal has a dominant high-frequency component, the first term in this formula will dominate the others. Maragos (1993) describes the secondary terms as "error" terms and shows how they can be minimized by constraining the AM and FM indexes of modulation, and the modulating signal bandwidth.

The integrals of the two terms in Teager's Ψ operator are both equal to the signal's total energy times its average square frequency; i.e.,

$$\int [s'(t)]^2\,dt = -\int s(t)s''(t),dt = \int \omega^2|S(\omega)|^2\,d\omega = E\langle\omega^2\rangle\,,$$

where $S(\omega)$ is the signal's Fourier transform, E is its total energy,

$$E = \int [s(t)]^2 dt = \int |S(\omega)|^2\,d\omega\,,$$

and $\langle\omega^2\rangle$ is the average square frequency.

Instantaneously, though, Teager's two terms are quite different. Ψ may not even yield positive results. For the signal in Fig. 4.1, for example, the values of Ψ are negative in the vicinity of $t = 2n\pi$ (where $s'(t) < 0$, $s(t) > 0$, and $s''(t) > 0$).

For lightly modulated signals, Ψ produces a stable output dominated by $[a(t)\varphi'(t)]^2$. Maragos (1992) showed that, as long as the "error" terms are sufficiently small, Ψ can be used to demodulate the signal and extract approximate values for $a(t)$ and $\varphi'(t)$ by applying Ψ to the signal

$$\Psi[s(t),t] = \Psi\{a(t)\cos[\varphi(t)],t\} \approx [a(t)\varphi'(t)]^2$$

and to its derivative

$$\Psi[s'(t),t] \approx a^2(t)[\varphi'(t)]^4\,.$$

Teager's formula appears to offer possibilities for identifying signals that would satisfy our general notion of monocomponentness. Turning these results into algorithms for separating monocomponent signals from more complex ones, however, is still an open problem.

We proceed from this point without a concrete definition of "monocomponentness," but we recognize that it implies constraints on phase monotonicity ($\varphi'(t) > 0$), amplitude and "carrier" signal bandwidth, and degrees of amplitude and frequency modulation.

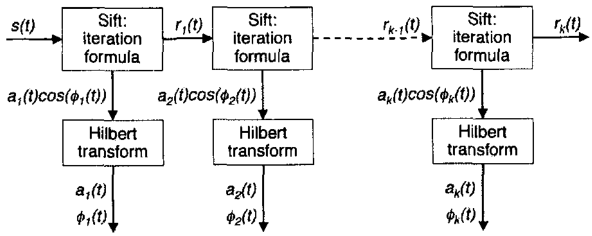

Figure 4.2: Diagram of the HHT signal separation process.

4.2.2. *End-point analysis*

An area of interest to many scientists is the extraction of frequency information at the very beginning and at the very end of the data they collect. We have not pursued this problem and are skeptical about prospects for significant advances in this area. Cohen (1995) states that the frequency content of a signal at any point in time depends entirely on the context of its behavior both before and after the time in question. In the absence of data providing this context, assumptions could be made about the signal's probable past and future behavior, but the analysis would then merely reflect these assumptions. The primary reason for our interest in time-frequency analysis techniques for transient and non-stationary signals is because their behavior is unpredictable. We conclude, therefore, that unless the necessary assumptions and predictions are strongly supported by additional knowledge about the physical processes underlying the signal, end-point analyses should be treated as at least somewhat suspect.

4.3. Huang's sifting algorithm

Huang's sifting process [see Huang (1998)] separates the highest-frequency component embedded in a multicomponent signal from all the lower-frequency components. The remaining lower-frequency components together make up the signal trend. A signal can be described in terms of its first component and residual trend functions by

$$s(t) = a_1(t)\cos[\varphi_1(t)] + r_1(t)\,.$$

The sifting process for a single component is repeated by using the trend output from one stage as the input for the next, producing the series of $a_i(t)\cos[\varphi_i(t)]$ terms that add up to reconstruct the original signal, $s(t)$. A diagram of this process is shown in Fig. 4.2.

To determine $r(t)$, Huang fit smooth envelope curves (using cubic splines) to the local maxima of the signal and to the local minima. The average of these two envelopes provides a rough estimate of $r(t)$. (Local maxima are referred to as "positive peaks" even though the signal values at those points may be positive or negative. Local minima are similarly referred to as "negative peaks.") Huang then applied an iteration scheme to refine the estimated trend. The iteration scheme can be formulated as

$$r_{(n+1)}(t) = r_{(n)}(t) + \rho(c_{(n)}, t)\,,$$

where $c_{(n)}(t) = s(t) - r_{(n)}(t)$. The function ρ represents the spline curve fitting and averaging process applied to the peaks of function $c_{(n)}(t)$. (Subscripts in parentheses indicate the iteration count.) This calculation is repeated (starting with $r_{(0)}(t) = 0$) until a fixed point is reached, and $\rho(c_{(n)}, t)$ converges to zero (within some small ϵ). Once the residual or trend function is determined, the difference between it and the input signal is the highest-frequency separated component, $c_i(t) = a_i(t)\cos[\varphi_i(t)]$.

To separate the $a_i(t)$ and $\varphi_i(t)$ functions, Huang computed the component's analytic signal by using Fourier transforms. The Fourier transform of a function's Hilbert transform satisfies the relation

$$\mathcal{F}\{\mathcal{H}[s(t)]\} = -i\, sign(\omega)\mathcal{F}[s(t)]\,,$$

where $\mathcal{F}(\cdot)$ is the Fourier transform and $\mathcal{H}(\cdot)$ is the Hilbert transform. The Fourier transform of a function's analytic signal can then be formulated as

$$\mathcal{F}\{\mathcal{A}[s(t)]\} = \mathcal{F}[s(t)] + sign(\omega)\mathcal{F}[s(t)]\,,$$

which is zero for all negative frequencies and double the input signal's values for all positive frequencies.

Taking a separated component's Fourier transform, zeroing its negative-frequency terms and doubling its positive-frequency terms, and then applying the inverse Fourier transform, produces the component's complex analytic signal. The magnitude of the analytic signal is $a(t)$ and the argument is $\varphi(t)$.

Huang called this separation technique "empirical mode decomposition," and the individual component signals "intrinsic mode functions." His colleagues later named the method "the Hilbert–Huang transform."

Components separated by this process are well behaved, although, because the process is defined only in terms of this algorithm, the mathematical monocomponentness properties they satisfy are not easily determined. The work described below represents an attempt to link the HHT results to filtering, the mathematical properties of which are well established.

4.4. Incremental, real-time HHT sifting

In Huang's original HHT algorithm, the data passed between the processing blocks in Fig. 4.2 are arrays containing an entire time series. The incremental algorithm [see Meeson (2002)] turns these batch-processing blocks into pipeline processes that operate incrementally on streams of data, passing one data sample at a time.

The first step in sifting is to identify signal peaks. The calculation of peak values and times in the incremental HHT algorithm is the same as in Huang's

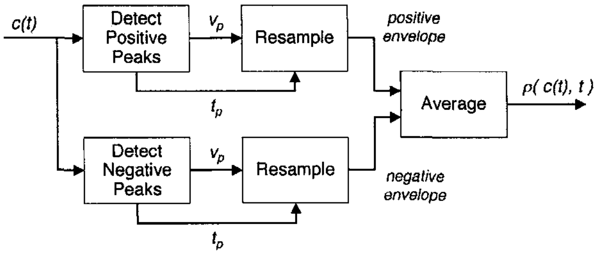

Figure 4.3: Diagram of one iteration of ρ.

original algorithm, except that peak value and time pairs, $\langle v_p, t_p \rangle$, are produced incrementally as the input stream evolves. The resulting stream of peak values corresponds to sampling the input signal at its peak times rather than at uniform intervals.

Spline interpolation uses global information to calculate the derivative of the positive envelope at each positive peak and for the negative envelope at each negative peak. For incremental processing, only local information is available, so we must rely on Hermite interpolation, which is nearly identical to spline interpolation but uses derivative values estimated from local signal behavior [see, for example, Kincaid (1991) for a thorough discussion of interpolation techniques].

By using the spline parameters derived for each segment of the positive peak envelope, values are calculated at points corresponding to the signal's original sample times. This resampling process produces a stream of uniformly sampled envelope values, although with some latency from the peak-detection and spline-interpolation process. The same process is applied to the negative-peak data. The two resampled envelope streams are then averaged to produce a stream of trend values. This process, diagrammed in Fig. 4.3, represents one application of Huang's ρ function. Each stage of this process is performed incrementally, so the calculation of ρ is achieved incrementally.

4.4.1. *Testing for iteration convergence*

The iteration process involving repeated applications of ρ does not always exhibit smooth convergence. Removing the residual or trend component occasionally exposes new peaks that appeared only as inflections in the input signal. An example where these peaks occur is illustrated by the signal

$$s(t) = \cos(t) - 0.167\cos(5t)\,.$$

The graph of this function is shown in Fig. 4.4. The trend function produced by ρ, also shown in the figure (dashed line), cuts through the inflection points in the signal as it crosses the axis and produces new peaks in the next iteration $c(t)$ that were not present in the input for this iteration. These new peaks are included in all further iterations.

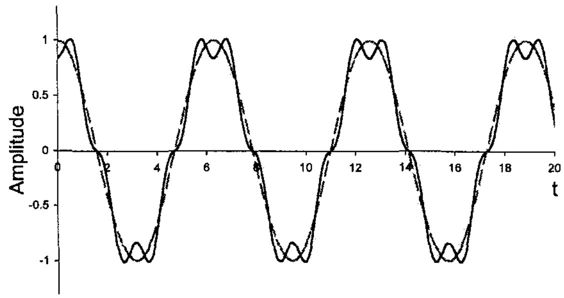

Figure 4.4: Example of a signal containing hidden peaks.

The discovery of new peaks introduces highly nonlinear disturbances in both the high-frequency component and the trend that may require several additional iterations to smooth out. These disturbances can occur even when the trend has nearly converged to a fixed point.

Huang devised a global test for convergence that spans the entire signal duration. This solution, however, is not in keeping with our objective of incremental processing. We have not yet found a satisfactory incremental test for convergence. In practice we have used fixed-length chains of ρ operations, making them long enough so that errors from terminating the iteration too early are rare. While not ideal for real-time performance, the results are comparable to those obtained by using the original HHT algorithm.

4.4.2. *Time-warp analysis*

If the peaks of an input signal are uniformly spaced, a number of simplifying assumptions can be made in the sifting process. These assumptions do not apply in general, so this approach cannot be used to process arbitrary signals, but the analysis provides insights that can be generalized.

Disregard, for the moment, the timing information that accompanies the incremental stream of peak values described above, and assume these peak values have been sampled at some uniform rate. The distortion this sampling introduces is referred to as a "time warp," since the actual peak times in general are not uniformly spaced. Although all of the nonlinear phase information between peaks in the original signal is lost (for the moment), the trend of the warped signal can be easily calculated by using standard low-pass digital filtering techniques.

In the warped world, one iteration of Huang's fixed-point function, ρ, for a series of warped peak values v at time t_p, corresponds to the expression

$$\rho(v, t_p) = \frac{v_p}{2} - \frac{v_{p-3}}{32} + \frac{9v_{p-1}}{32} + \frac{9v_{p+1}}{32} - \frac{v_{p+3}}{32}\,.$$

This expression is the average of the two envelopes, one of which is represented by v_p, and the other is interpolated from the spline curve derived from the neighboring

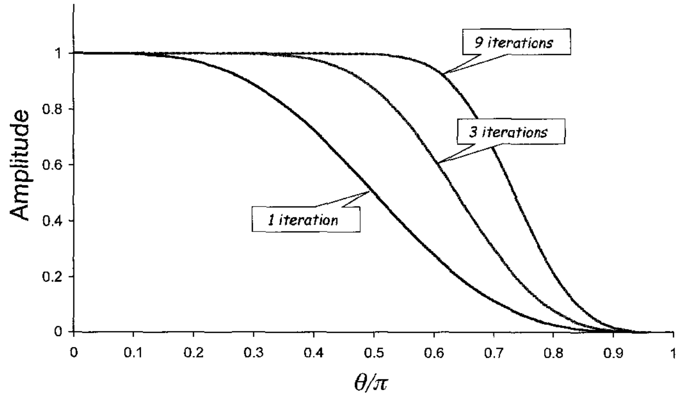

Figure 4.5: Frequency response of HHT trend-estimating process.

opposite-sign peaks (v_{p-3}, v_{p-1}, v_{p+1} and v_{p+3}) at time t_p. This expression corresponds to a simple low-pass digital filter, which has the frequency response shown in Fig. 4.5. This graph shows that the transition band for one pass through this filter crosses at approximately one-half of the warped signal's Nyquist frequency.

If the timing of peaks does not change from iteration to iteration, multiple iterations correspond to passing the signal through this filter multiple times. (The timing of peaks may change slightly, usually in the initial iterations.) Multiple passes through a simple filter are equivalent to a single pass through a larger filter [see, for example, Parks (1987) for a discussion of filter composition]. Huang's iteration scheme is formulated so that the high-pass filter is iterated and its iteration successively reduces the resulting pass band. The corresponding longer low-pass filters have wider pass-band regions and sharper transitions to the stop band. Examples of transfer functions for filters representing different iterations of ρ are shown in Fig. 4.5.

The original HHT algorithm uses the shrinking corrections of the iteration process to judge when it has converged. This method corresponds to choosing the characteristics of filters dynamically, based on the signal's behavior. Figure 4.5 shows that each iteration of ρ shifts the filter transfer function to a higher-frequency cutoff point. Note also that successive iterations have less and less effect on the size of the frequency shift. Rather than iterate the simple filter corresponding to ρ, we wish to determine the filter characteristics necessary to directly satisfy the monocomponent criteria and separate the component from the trend in a single pass.

4.4.3. *Calculating warped filter characteristics*

Consider that the separated warped signal can be described by $s_p = a_p + r_p$ for all positive peaks, and $s_p = -a_p + r_p$ for all negative peaks, where a_p is the absolute value of the high-pass filter output, and r_p is the low-pass filter output for each peak. The a_p values are interpreted as approximating a warped sampling of the amplitude

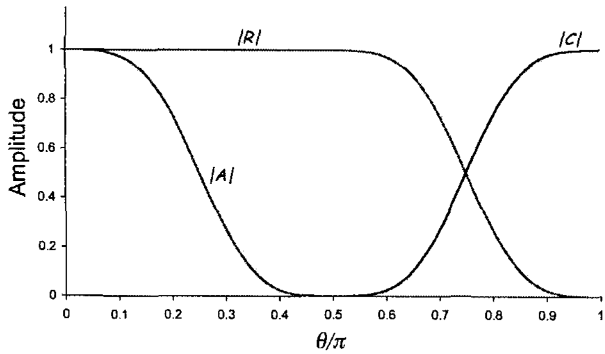

Figure 4.6: Warped filter transfer functions.

function, $a(t)$. The r_p values are similarly interpreted as a warped sampling of the residual function, $r(t)$.

The spectrum of the warped residual function is controlled by the low-pass filtering effects of multiple iterations of ρ. This same filtering process also controls the spectrum of the warped amplitude function. The spectrum of the series of a_p values is shifted upward by the modulating effects of the warped "carrier" signal, $\cos(p\pi)$. The spectrum captured by the high-pass filter, therefore, is that of the amplitude function shifted upward by π. If $R(\theta)$ is the low-pass filter transfer function for the r_p values, then the corresponding transfer function for the a_p values is

$$A(\theta) = 1 - R(\pi - \theta)\,.$$

This relationship, for an idealized separation filter, is shown in Fig. 4.6. [The transfer function for the high-pass filter is shown as $C(\theta)$.] From these graphs we can see that, to satisfy Bedrosian and keep the $\cos(p\pi)$ and a_p spectra from overlapping, the stop band breakpoint for the high-pass filter must be no lower than half the warped Nyquist frequency.

Ten iterations of this filter would reduce the effective filter throughput at one-half the warped Nyquist frequency to approximately $2^{-10}(\approx 10^{-3})$, which should satisfy Bedrosian's separation criteria for many practical purposes. Iterating the simple warped filter or substituting a more efficient filter, however, will not adapt to the frequency changes introduced by new peaks. In practice, we have often encountered signals that require 25 to 30 iterations of Huang's ρ operator to converge. Much of this disparity in iteration counts is attributable to the nonlinear disturbances caused by the discovery of new peaks.

4.4.4. *Separating amplitude and phase*

To separate the amplitude and phase functions incrementally, we substituted a Hilbert transform filter for the batch Fourier transform process described earlier for calculating analytic signals. A Hilbert transform filter has a transfer function that approximates the Fourier transform of a signal's Hilbert transform: $H(\omega) = -jsign(\omega)$ [see, for example, Parks (1987) for a discussion of Hilbert transformer filter design]. For a monocomponent signal, $a(t)\cos[\varphi(t)]$, this filter approximates

$$h(t) * \{a(t)\cos[\varphi(t)]\} = a(t)\sin[\varphi(t)]\,,$$

where $h(t)$ represents the Hilbert transform filter coefficients and "$*$" represents convolution. The amplitude and phase functions are easily separated by using this result:

$$a(t) = \sqrt{\{a(t)\sin[\varphi(t)]\}^2 + \{a(t)\cos[\varphi(t)]\}^2}\,,$$

and

$$\varphi(t) = \arctan\{\sin[\varphi(t)]/\cos[\varphi(t)]\}\,.$$

Once the phase function is extracted, the signal's instantaneous frequency is calculated by passing $\varphi(t)$ through a differentiating filter (after compensating for the discontinuities in the arctan results). All of these calculations are done incrementally.

The band-limiting effects of warped filtering on the amplitude envelope indicate that $a(t)$ should be relatively smooth. That is, we expected $a(t)$ to look like the smooth spline-connected envelopes calculated in the final iteration of ρ in the sifting process, with all of the high-frequency content captured by the phase function, $\varphi(t)$. Both the Hilbert transform filter and the Fourier batch technique, however, were found to introduce a high-frequency "ripple" in the amplitude results for some signals.

The explanation for this seeming anomaly is that, within certain limits, the spectral energy of a combined amplitude- and frequency-modulated signal can be freely exchanged between the amplitude and phase functions. While we expected a band-limited amplitude, the Hilbert transform appears to split the difference, sharing the high-frequency content between the amplitude and phase functions. The result, therefore, is sometimes a bit different from what we expected, but is an equivalent representation of the signal.

We experimented with a number of different possible techniques for separating the amplitude and phase, including the use of Teager's energy operator. None of these other techniques were as successful as the use of the Hilbert transform filter. Teager's operator worked well for the signal itself, but occasionally produced negative results for the derivative of the signal, spoiling Maragos's (1992) demodulation approach. Boashash (1992) provides an extensive discussion of additional techniques for extracting a signal's instantaneous frequency.

4.5. Filtering in standard time

The next objective, to test our conjecture about filtering substituting for HHT sifting, was to reproduce the effects of Huang's ρ operation in standard time, with-

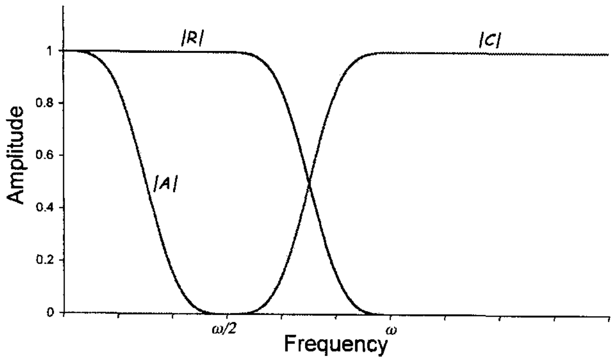

Figure 4.7: Filter transfer functions.

out resampling the original input signal. In the process, we wanted to avoid the unreasonable time-warp analysis assumptions about uniformly spaced peaks. The question posed was, Is there a corresponding standard-time filter that will isolate a comparable (unwarped) trend function and, if so, what are its characteristics? Any filter that approximates this response will have to change its attributes over time (possibly every few samples) to track transient and non-stationary changes in the signal.

The transfer function for this low-pass filter is shown schematically in Fig. 4.7 as $R(\theta)$. The transfer function for the complementary high-pass filter for the $a(t)\cos[\varphi(t)]$ term is shown as $C(\theta)$. This filtering should also leave the spectrum of the amplitude function as shown by $A(\theta)$, maintaining Bedrosian's separation from the minimum frequency of the $\cos[\varphi(t)]$ term. All we have to do is determine the breakpoint frequencies, ω and $\omega/2$, for these filters and calibrate the horizontal scale.

The spectrum of the $a(t)\cos[\varphi(t)]$ term will, in general, contain both AM and FM components. Amplitude modulation of a constant-frequency "carrier" signal shifts the spectrum of the amplitude signal from the origin to the carrier frequency. If $A(\theta)$ is the spectrum of $a(t)$, then the spectrum of $a(t)\cos(\omega_c t)$ will be $A(\theta + \omega_c)$, where ω_c is the carrier frequency. Frequency modulation redistributes the spectrum of its modulating signal in much more complex ways.

In a combined AM and FM signal, the FM spectrum overlaps and mixes with the AM spectrum so that separating the two components by using a simple linear process (like conventional filtering) does not appear promising. The HHT process, however, is able to make a separation, although not always in exactly the same form as that used to formulate sample inputs. (Recall that solutions satisfying the HHT monocomponent separation criteria are not unique.)

As a first approximation for the breakpoint for the high-pass filter pass band, we used the minimum peak-to-peak frequency of the signal over the time span covered by the filter.§ This frequency is marked as ω on the axis in Fig. 4.7. The pass band breakpoint for the high-pass filter was set to this frequency. The stop band breakpoint, based on our experience with warped filtering, was set to one-half this frequency. As signals pass through the filter, their peak-to-peak frequencies are monitored, and the filter coefficients are adjusted to track any changes.

4.6. Case studies

In this section we present five case studies that illustrate and compare the results produced by the HHT and filtering approaches. The first example is a simple composite signal that serves as a reference for comparison with the second example. The second example is a steady-state AM signal. The third example is a steady-state FM signal. The fourth and fifth examples contain unit step changes in amplitude and frequency, respectively, and allow us to begin to explore the dynamic capabilities of the HHT and filtering mechanisms.

4.6.1. *Simple reference example*

The first example is a simple combination of constant amplitude sinusoids defined by

$$s(t) = \cos(t) + 0.5\cos(t/2)\,.$$

The graph of this function is shown in Fig. 4.8, along with the signal trend (dotted line). The maximum timing between peaks is slightly greater than π, indicating the need for high- and low-pass filters with upper breakpoint frequencies at $\omega = 0.97$. The result of filtering this signal, because of our selection of filter breakpoints, produces a nearly perfect separation of the two components, namely $c_1(t) = \cos(t)$

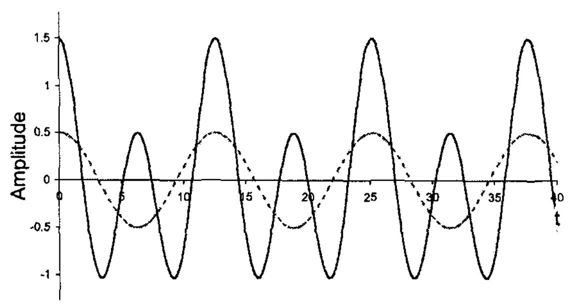

Figure 4.8: Simple two-component example signal.

§We note that Bedrosian's spectral separation criteria, being based on integral transform analysis, must hold (theoretically) for all time, not merely for the time span covered by the filter. We conjecture that this rather severe constraint can be relaxed by using more modern tight-frame analysis techniques. We have not completed the analysis to formally confirm this conjecture, however, and we proceed, taking it as an assumption.

and $r_1(t) = 0.5\cos(t/2)$. The HHT sifting process produces nearly identical results. One difference is that HHT sifting approximates the trend using splines, so its trend is represented by a series of cubic polynomials pieced together at the peaks. These small differences are of little concern here. Our primary interest in this simple signal is its similarity to the next example.

4.6.2. *Amplitude modulated example*

The second example is a stationary amplitude-modulated signal defined by

$$s(t) = [1 + 0.5\cos(t/2)]\cos(t)\,.$$

The graph of this function is shown in Fig. 4.9, along with this function's positive and negative envelope (dotted lines). Note that a very similar envelope could also be constructed for the previous example.

The differences between this example and the first one are that the tall positive peaks are a little narrower, and the shorter positive peaks are a little broader. The positive peaks have exactly the same values and timing. The negative peaks extend slightly lower (to -1.03), and their timing is shifted slightly toward the tall positive peaks. Another way to examine these signals is to expand this example's definition and apply a trigonometric identity for the product of two cosines:

$$\begin{aligned} s(t) &= \cos(t) + 0.5\cos(t/2)\cos(t) \\ &= \cos(t) + 0.25\cos(t/2) + 0.25\cos(3t/2)\,. \end{aligned}$$

The above equations show that the difference between this example and the previous one is a smaller coefficient for the $\cos(t/2)$ term and an additional higher-frequency term, $0.25\cos(3t/2)$.

The maximum timing between peaks is again slightly greater than π, indicating the need for filters with upper breakpoint frequencies at $\omega = 0.93$. The result of filtering this signal separates the lower-frequency $\cos(t/2)$ term from the two higher frequency components; i.e.,

$$c_{filter}(t) = \cos(t) + 0.25\cos(3t/2)$$

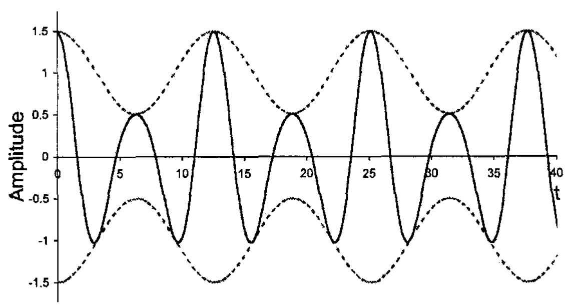

Figure 4.9: Example of an amplitude-modulated signal.

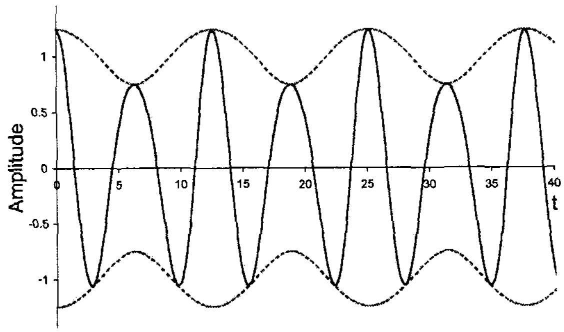

Figure 4.10: High-frequency component separated from the AM signal by filtering.

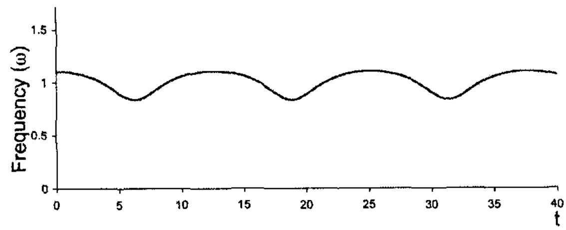

Figure 4.11: Instantaneous frequency of the AM signal component separated by filtering.

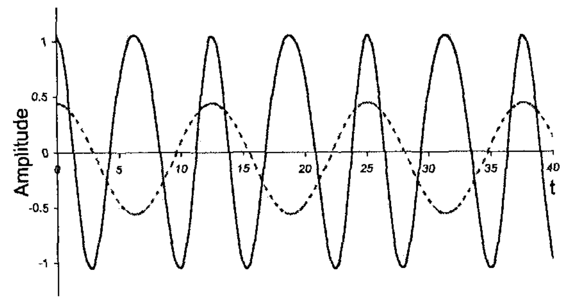

Figure 4.12: *High-frequency component separated from the AM signal by the HHT sifting process.*

and

$$r_{filter}(t) = 0.25\cos(t/2)\,.$$

The high-frequency component produced by filtering, $c_{filter}(t)$, is shown in Fig. 4.10, along with its amplitude envelope. The instantaneous frequency of the filtering solution ranges from approximately 0.83 to 1.10, as shown in Fig. 4.11.

The result produced by HHT sifting is quite different, as shown in Fig. 4.12. The HHT sifting algorithm produces a constant-amplitude, frequency-modulated component, and a trend that is the same frequency as the filter trend but twice its amplitude. The component's constant amplitude makes its frequency modulation more clearly evident.

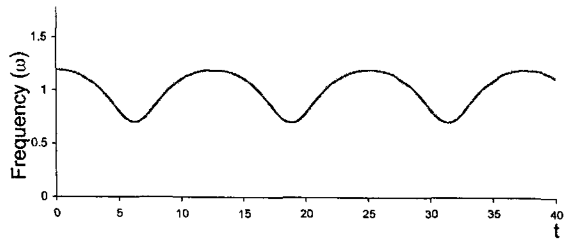

Figure 4.13: Instantaneous frequency of the AM signal component separated by HHT sifting.

The HHT-separated component and trend signals can be described mathematically by

$$c_{HHT}(t) = \cos(t) + 0.25\cos(3t/2) - 0.25\cos(t/2) + 0.0563$$

and

$$r_{HHT}(t) = 0.5\cos(t/2) - 0.0563\,.$$

The small constant terms in the HHT formulas offset the frequency modulation effects that result when the three cosine terms in $c_{HHT}(t)$ are combined. These effects are discussed in the next example.

The instantaneous frequency of this signal, shown in Fig. 4.13, has a larger range than that for the filter solution. The instantaneous frequency of the HHT sifting solution ranges from approximately 0.69 to 1.19.

Both solutions produce monocomponent high-pass components and band-limited trend signals, satisfying the HHT objectives as we characterized them earlier. The filter produces a mixed AM and FM component with a smaller-amplitude trend signal. HHT sifting produces a purely FM component with larger frequency variations and a larger-amplitude trend signal.

In this example, the HHT result illustrates a classic example of signal aliasing. The HHT and warped filtering processes, being based on peak values, which are sampled at a frequency of $\omega = 2$, under-sample the input signal and misinterpret the energy from the higher-frequency $\cos(3t/2)$ component and attribute it incorrectly to the lower-frequency $\cos(t/2)$ term. The extra $\cos(t/2)$ energy in both the HHT component and the trend for this signal does not accurately redistribute the $\cos(t/2)$ energy contained in the input signal. Aliasing often leads to unintended consequences, which we believe is the case here.

4.6.3. *Frequency modulated example*

The third example is a stationary frequency-modulated signal defined by

$$s(t) = \cos[t + 0.5\sin(t)]\,.$$

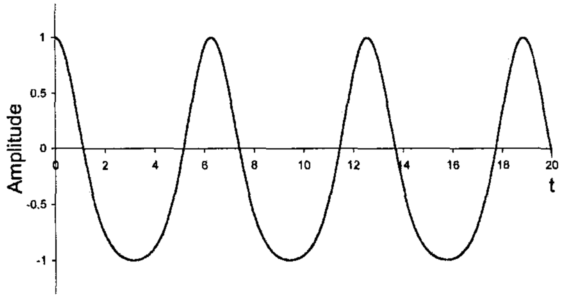

Figure 4.14: Example of a frequency-modulated signal.

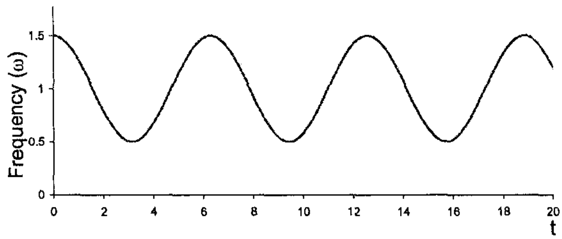

Figure 4.15: Instantaneous frequency of the FM signal component separated by HHT sifting.

The amplitude of this signal is constant, but its phase increases nonlinearly. The graph of this function, shown in Fig. 4.14, shows sharpened positive peaks and rounded negative peaks, much like the solutions to Stokes's equation (although these results are not a solution to Stokes's equation). See Huang (1999) for further discussion and analysis of nonlinear wave dynamics.

HHT analysis of this signal finds evenly spaced constant-valued positive and negative peaks. The trend function is a constant zero, and the separated component captures the entire signal. The instantaneous frequency derived from the HHT results, as shown in Fig. 4.15, matches our expectations: $\varphi'(t) = 1 + 0.5\cos(t)$.

The filtering results are a bit more complicated to explain. The coefficients of the Fourier series for a frequency-modulated signal are defined in terms of Bessel functions. [See, for example, Lathi (1965) or Schwartz (1990) for details.] If the signal is generalized to

$$s(t) = A\cos[\omega_c t + \beta\sin(\omega_m t)]\,,$$

where A denotes the signal's constant amplitude, ω_c is its "carrier" frequency, β is the index of modulation, and ω_m is the modulating frequency, then the equivalent Fourier series is

$$s(t) = A\sum_n J_n(\beta)\cos[(\omega_c + n\omega_m)t]\,,$$

where $J_n(\cdot)$ is the Bessel function (first kind) of order n. The summation, theoreti-

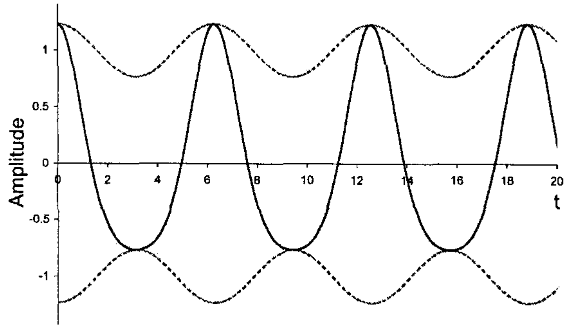

Figure 4.16: High-frequency component separated from the FM signal by filtering.

cally, ranges over integral values of n from $-\infty$ to ∞. Bessel function values for small values of β, however, are essentially zero for all but a few terms. An approximate Fourier series for this signal is

$$s(t) \approx -0.242 + 0.969\cos(t) + 0.242\cos(2t) + 0.031\cos(3t)\,.$$

This representation of the signal shows that its nonlinear phase gives it a constant "DC" term as well as higher-frequency harmonic components. The filter breakpoint frequencies for this signal, determined from the signal's peak-to-peak timing, were $\omega = 1$ and $\omega = \frac{1}{2}$. This filter produced the separation

$$c_{filter}(t) = 0.969\cos(t) + 0.242\cos(2t) + 0.031\cos(3t)\,,$$

and

$$r_{filter}(t) = -0.242.$$

The results for the high-pass component are shown in Fig. 4.16, along with a smooth amplitude envelope connecting the absolute values of the peaks (dashed lines).

The output from this filter differs from the monocomponent signal we started out with, although the basic shape of the input signal is preserved. The oscillating amplitude appears problematic, since the input signal contained no amplitude modulation. Furthermore, the amplitude oscillations have the same average frequency as the signal, which violates Bedrosian's spectral separation conditions. These amplitude oscillations appeared because the filtering process removes the constant term in the signal's Fourier series. Our earlier time-warp analysis showed that the amplitude envelope should be band-limited to below one-half of the signal's "carrier" frequency. The observed higher-frequency content, therefore, is an unexpected artifact that must be attributed to the filtering process.

Similar nonlinear signal behavior was encountered in the previous (AM) example. The high-frequency component separated by HHT sifting (shown in Fig. 4.12) contains alternating narrow and wide positive peaks. This nonlinear phase behavior gives this signal a constant term similar to that described here. As these examples

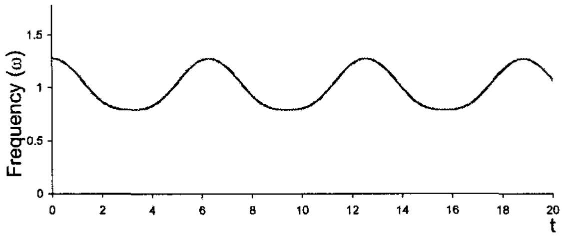

Figure 4.17: Instantaneous frequency of the FM signal component separated by filtering.

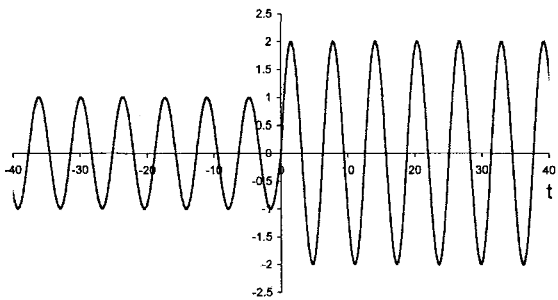

Figure 4.18: Amplitude step example signal.

show, any signals with nonlinear phase behavior may contain low-frequency energy that will produce similar artifacts in filter results.

The instantaneous frequency derived from the high-pass filter output is shown in Fig. 4.17. This signal has a smaller frequency range than the HHT component ($\omega = 0.79$ to 1.28), and the variations are not purely sinusoidal.

4.6.4. *Amplitude step example*

The preceding examples are all stationary signals that could be handled by static filtering techniques (if the frequencies are known in advance). The signal shown in Fig. 4.18 begins to exercise the dynamic capabilities of the HHT and filtering processes. This signal contains a step discontinuity in its amplitude at time $t = 0$; i.e.,

$$s(t) = \begin{cases} \sin(t) & \text{for } t \leq 0 \\ 2\sin(t) & \text{for } t \geq 0\,. \end{cases}$$

Both the HHT and filtering processes are expected to smooth out this amplitude transition because of limitations on amplitude bandwidth suggested by the mono-componentness considerations. The results plotted in Figs. 4.19 and 4.20 verify this expectation. The differences in smoothing are a result of the differing filter transfer functions and, in the case of the HHT, its signal aliasing behavior. The trend signals

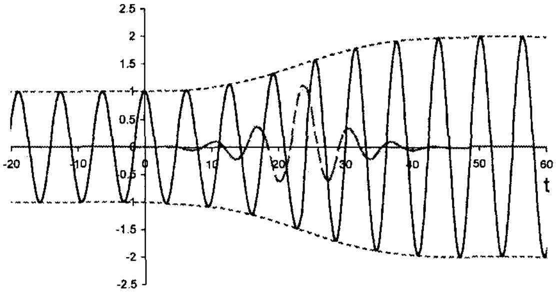

Figure 4.19: HHT component and trend results for the amplitude step signal.

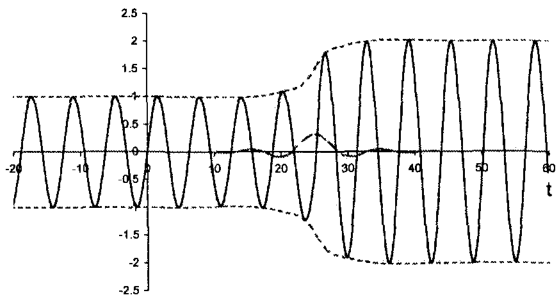

Figure 4.20: Filter high- and low-pass results for the amplitude step signal.

in both cases are shaped somewhat like sampling functions. The HHT trend has a considerably higher amplitude than the filter low-pass signal.

There is also a time delay of approximately 24 time units for the incremental HHT result and 25 time units for the filtering results. These delays are necessary to collect data on the signal's future behavior, which both processes need before they can produce their results.

The instantaneous frequencies, derived numerically, for the two separated high-pass components are shown in Figs. 4.21 and 4.22. In both cases, the effect of smoothing out the amplitude step transient has created transient frequency modulations. This result suggests the presence of a "conservation of transient energy" law that allows amplitude transients to be transformed into frequency transients.

Although our understanding of this frequency behavior is incomplete, we can explain the behavior of the two signal separation processes by using their representation in the frequency domain. The Fourier transform of the amplitude step signal is

$$S(\omega) = 3\pi i[\delta(\omega + 1) - \delta(\omega - 1)]/2 + 1/(1 - \omega^2)\,,$$

where $\delta(\cdot)$ denotes the Dirac delta function. Figure 4.23 shows the magnitude of this transform. It has complex poles at $\omega = \pm 1$, which reflect the $\sin(t)$ term in the

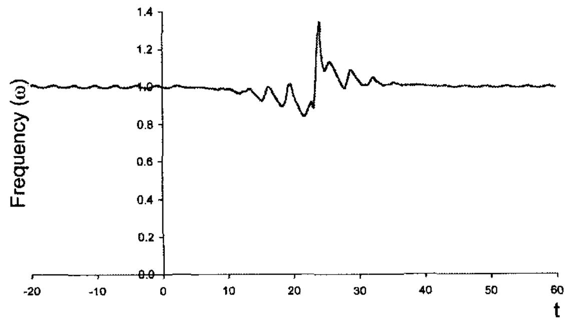

Figure 4.21: Instantaneous frequency of the amplitude step component separated by HHT sifting.

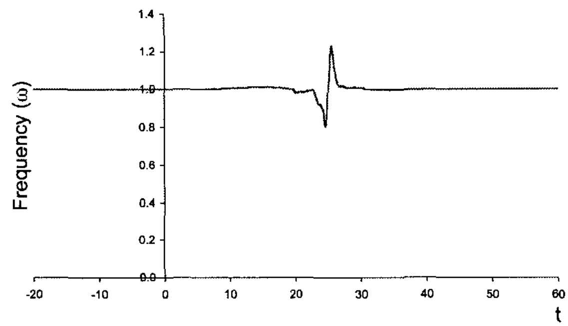

Figure 4.22: Instantaneous frequency of the amplitude step component separated by filtering.

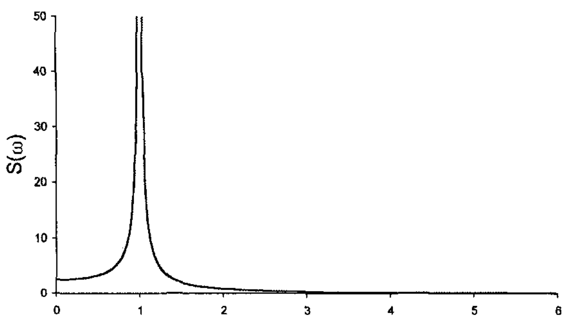

Figure 4.23: Fourier transform (magnitude) of the amplitude step signal.

signal. The bandwidth contributed by the amplitude step is distributed smoothly over the entire frequency spectrum.

Figure 4.24 shows how filtering separates the amplitude step signal in the frequency domain. The low-pass (solid) curve shows the spectrum of the signal trend and the high-pass results (dashed) curve shows the spectrum of the separated com-

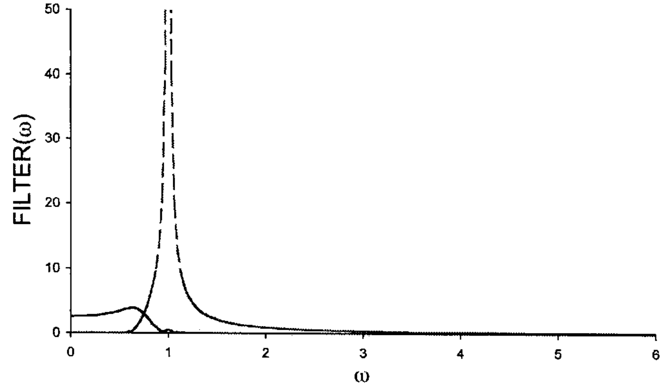

Figure 4.24: Filter high- and low-pass spectra for the amplitude step signal.

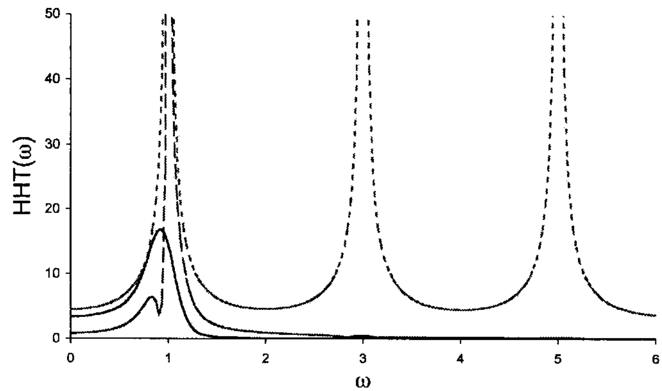

Figure 4.25: Spectra of the HHT trend and separated component for the amplitude step signal.

ponent. Inverting these transforms back into the time domain reproduces the trend and component signals shown in Fig. 4.20.¶

In practice, the results shown back in Fig. 4.20 are produced by direct convolution of the signal with the digital filter coefficients, not by applying transforms. The results, however, are the same by using either process.

An explanation of the HHT results requires introducing the effects of the peak curve-fitting and iteration process. Figure 4.25 shows the spectra of the signals separated by the HHT algorithm. The low-frequency "hump" (solid line) is the

¶Care must be taken with numerical Fast Fourier Transform (FFT) tools when analyzing these signals and spectra. The results presented here are for continuous infinite-integral transforms of one-time transient events. Numerical techniques that operate on finite-duration numerical representations of signals and their spectra can easily generate different results. For example, a finite representation of the signal shown in Fig. 4.18 will be presumed to repeat periodically. While the graph still looks like a one-time unit step amplitude change, the transform produced will be for a repeating square-wave modulated signal.

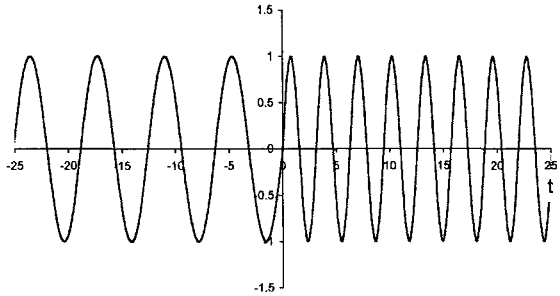

Figure 4.26: Frequency shift example signal.

trend's spectrum. The second curve (dashed line) is the spectrum of the separated high-frequency component. Transforming these spectra back into the time domain reproduces the signal trend and separated component shown in Fig. 4.19.

The third curve in Fig. 4.25 (dotted line) shows the apparent spectrum of the signal that was derived by resampling it at its peaks. This result is a direct effect of aliasing. Because the peak sampling rate is below the signal's original sampling rate, aliasing creates overlapping replicas of the spectrum shown in Fig. 4.23. The results shown in Fig. 4.25 reveal that aliasing has a significant effect on the signal's apparent spectrum, causing the HHT algorithm to attribute considerable energy to the trend that is not part of the input signal. The separated high-frequency component is calculated by resampling the trend at its original sample times and subtracting that result from the original input signal. Only the trend, therefore, is directly affected by the aliased spectrum.

4.6.5. *Frequency shift example*

The final example signal to be explored contains a step discontinuity in frequency at time $t = 0$, given by

$$s(t) = \begin{cases} \sin(t) & \text{for } t \leq 0 \\ \sin(2t) & \text{for } t \geq 0. \end{cases}$$

A graph of this signal is shown in Fig. 4.26. Because the signal amplitude is constant, the HHT trend remains constant (zero) through this frequency shift. The aliasing has no effect because the trend is zero. The first separated HHT component captures the entire input signal. It seems clear from this example and the earlier frequency-modulated example that the HHT will separate any constant-amplitude, monotonically-increasing phase signal as a single component.

The instantaneous frequency extracted from the signal, which is the HHT-separated component, is shown in Fig. 4.27. It tracks the signal nearly perfectly through the transition. While the HHT produced considerable smoothing of the amplitude step in the previous example, it makes no attempt to smooth out the frequency shift here.

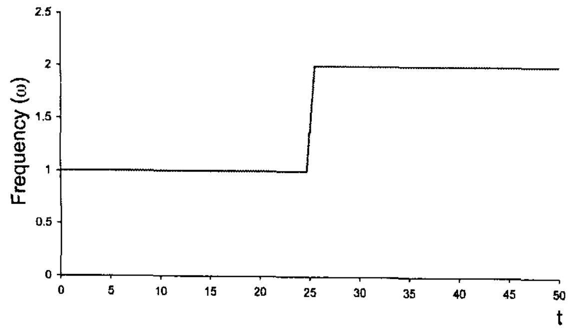

Figure 4.27: Instantaneous frequency of the frequency shift component separated by HHT sifting.

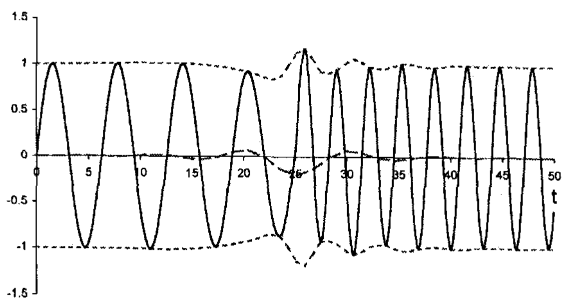

Figure 4.28: Filter high- and low-pass results for the frequency shift signal.

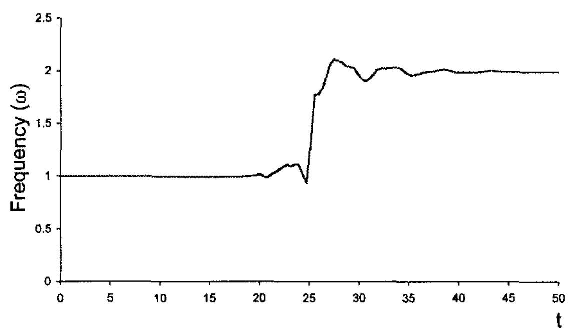

Figure 4.29: Instantaneous frequency of the frequency shift component separated by filtering.

Filtering produces quite different results, as shown in Fig. 4.28. The high-pass signal (solid line) shows a clear disturbance, although it is difficult to characterize. The low-pass signal (central dotted line) looks something like an inverted sampling function, centered at the point where the frequency shift takes place. The amplitude envelope around the high-frequency signal (upper and lower dotted lines) also reflects the disturbance.

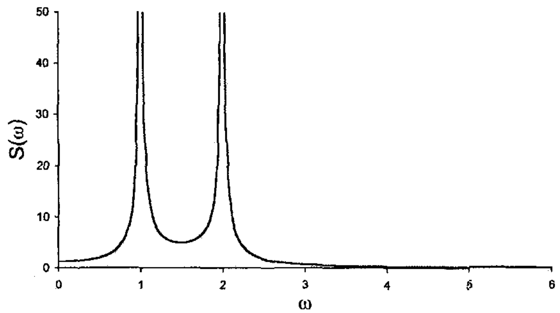

Figure 4.30: Fourier transform (magnitude) of the frequency shift signal.

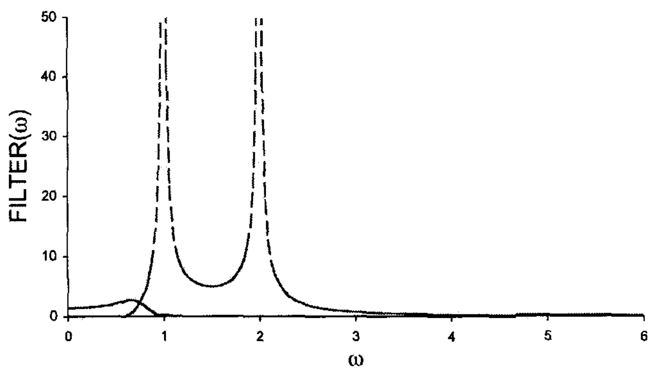

Figure 4.31: Filter high- and low-pass spectra for the amplitude shift signal.

The instantaneous frequency, derived numerically, for the high-pass component signal is shown in Fig. 4.29. We do not fully understand why the frequency behavior should take this shape.

As with the previous example, we turn to the frequency domain to explain the behavior of the filter. The Fourier transform of the frequency step signal is

$$\begin{aligned} S(\omega) = \ & \pi i[\delta(\omega+1) - \delta(\omega-1)]/2 + \pi i[\delta(\omega+2) - \delta(\omega-2)]/2 \\ & -1/(1-\omega^2) + 2/(4-\omega^2)\,. \end{aligned}$$

The magnitude of this transform is shown in Fig. 4.30. The complex poles at $\omega = \pm 1$ and $\omega = \pm 2$ reflect the signal's two sinusoidal frequencies. The bandwidth contributed by the frequency transition is distributed smoothly over the entire spectrum.

Figure 4.31 shows how filtering separates the frequency shift signal in the frequency domain. The low-pass curve (solid line) shows the spectrum of the signal trend. The high-pass curve (dashed line) shows the spectrum of the separated component. Inverting these transforms reconstructs the signals shown in Fig. 4.28. The results shown in Fig. 4.28, however, were calculated by direct convolution of the signal with the digital filter coefficients, not by applying transforms.

Because the filter breakpoint frequencies are determined by the lowest peak-to-peak frequency within the span of the filter, these results are effectively the same as for static filters with breakpoints at $\omega = \frac{1}{2}$ and $\omega = 1$. Once the last low-frequency peak passes through the filter, its coefficients are adjusted to move the breakpoint frequencies to $\omega = 1$ and $\omega = 2$. This adjustment has no effect on the filter outputs because, in both cases, the signal resides completely within the high-pass pass band.

4.7. Summary and conclusions

In this chapter, the HHT component separation, or "sifting," process has been compared with a filtering process that was intended to mimic HHT behavior. The conjecture that conventional digital filters, by adapting dynamically to signal frequency content, could substitute for the HHT process was found to be incorrect. The results from several example signals showed that under most conditions, the two techniques produce distinct results.

The experiments we conducted to compare the HHT and filtering processes led to the discovery of aliasing in the HHT sifting algorithm. The process of sampling a signal at its peak times results in a classic example of under-sampling that leads to misinterpretation of signal frequency content. Specifically, signal content at frequencies above the peak-to-peak sampling rate is misinterpreted as lower-frequency content.

The question of whether aliasing is a problem or a "feature" in terms of HHT signal separations has not yet been completely resolved. High-frequency components separated from examples where aliasing arises appear to satisfy the requirements for "monocomponentness," so they are expected to have well-defined instantaneous frequencies. Where aliasing arises, however, it introduces anomalous energy into both the high-frequency component and the trend, and this result can be considered a form of signal corruption. Further investigation is needed to determine if unaliased filtering results are indeed "better," or if the aliasing is in some unusual way a necessary aspect of the HHT sifting process.

4.7.1. *Summary of case study findings*

For signals with a dominant highest frequency (case study #1), the HHT and filtering were found to produce equivalent separations.

For stationary amplitude-modulated signals with a dominant central "carrier" frequency (case study #2), filtering separates the lower sidebands as the trend, and the carrier and upper sidebands as the high-frequency component. The HHT, because of aliasing, misinterprets the upper sideband energy as lower-frequency energy, effectively doubling the lower sideband amplitude. This result gives the high-frequency component a nearly constant amplitude and larger variations in instantaneous frequency.

For signals with transient amplitude changes (case study #4), HHT sifting produced a broad smoothing of the amplitude transition and, because of aliasing, a large trend amplitude. Filtering also smoothed out the amplitude transition, but not as broadly as the HHT. Its trend amplitude was small compared to that of the HHT trend.

For frequency-modulated signals with monotonically increasing phase (case studies #3 and #5), the HHT high-frequency component captures the entire signal, leaving a zero-valued residual trend. The extracted phase function, $\varphi(t)$, and instantaneous frequency, $\varphi'(t)$, for these signals tracked the signal behavior very closely, even with significant transients in frequency (case study #5). Filtering had considerably more difficulty with FM signals. Signals with nonlinear phase functions often have significant low-frequency content. Conventional filtering separates the high- and low-frequency energy, disrupting the input signal's monocomponent characteristics.

4.7.2. *Research directions*

Although this paper investigated a key step in separating signal components, additional aspects of the overall problem need attention. The following research areas have been identified as areas still to be explored.

Resolving the question about aliasing is of high priority. Our preference for a solution would be an algorithm that separates complex signals into components without aliasing, and without the amplitude disturbances filtering causes with FM signals.

Second on our list is finding a better way to separate amplitude and phase information from monocomponent signals. Although the Hilbert transform is the obvious theoretical solution, the finite numerical approximations we used produced anomalous results.

Episodes of signals with only very low-frequency content compared to their sampling rate (that is, with many samples between peaks) would require excessively long filters to achieve the separations we propose. To process such signals, a method is needed for adaptively down-sampling or decimating the signal and for automatically restoring higher sampling rates when higher-frequency content returns. Static down-sampling is used extensively in wavelet transform processing [see, for example, Daubechies (1992)]. To our knowledge, the idea of a dynamic down-sampling mechanism has not been explored.

The residual trend signals that are passed to successive stages of sifting have their high-frequency content removed, resulting in signals with lower and lower frequency content. This result is a prime example of where signal down sampling is needed. Non-uniform sampling techniques may be useful here, although they appear to require more complex up-sampling procedures to restore their original sampling rates than do uniformly sampled signals [see Aldroubi (2001) for examples of possible techniques].

Real-world signals often contain components that turn on and off intermittently, like the telephone that rings while we are listening to our favorite music or are eating dinner. Huang (1999) developed a technique for dealing with such intermittent components that attempts to minimize the disturbance in analysis of more continuous "background" components. Although a clear need exists for this capability, it has not yet been addressed in our real-time algorithms.

References

Aldroubi, A., and K. Gröchenig, 2001: Nonuniform sampling and reconstruction in shift-invariant spaces. *SIAM Review*, **43**, 585–620.

Bedrosian, E., 1963: A product theorem for Hilbert transforms. *Proc. IEEE*, **51**, 868–869.

Boashash, B., 1992: Estimating and interpreting the instantaneous frequency of a signal. *Proc. IEEE*, **80**, 520–568.

Cohen, L., 1995: *Time-Frequency Analysis*. Prentice Hall, 299 pp.

Daubechies, I., 1992: *Ten Lectures on Wavelets*. CBMS-NSF Series in Applied Mathematics, Vol. 61, SIAM, 357 pp.

Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London, Ser. A*, **454**, 903–995.

Huang, N. E., Z. Shen, and S. R. Long, 1999: A new view of nonlinear water waves: The Hilbert spectrum. *Annu. Rev. Fluid Mech.*, **31**, 417–457.

Huang, N. E., C. C. Chern, K. Huang, L. W. Salvino, S. R. Long, and K. K. Fan, 2001: A new spectral representation for earthquake data: Hilbert spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999. *Bull. Seism. Soc. Am.*, **91**, 1310–1338.

Kincade, D., and W. Cheney, 1991: *Numerical Analysis*. Brooks/Cole Publ., 690 pp.

Lathi, B. P., 1965: *Signals, Systems and Communication*. John Wiley & Sons, 607 pp.

Maragos, P., J. F. Kaiser, and T. F. Quatieri, 1992: On separating amplitude from frequency modulations using energy operators. Preprints, *Intl. Conf. on Acoust., Speech, and Signal Process.*, San Francisco, CA, IEEE, Vol. 2, 1–4.

Maragos, P., J. F. Kaiser, and T. F. Quatieri, 1993: On amplitude and frequency demodulation using energy operators. *IEEE Trans. Signal Process.*, **41**, 1532–1550.

Meeson, R., 2002: An Incremental, Real-Time Algorithm for the Hilbert Huang Transform. IDA Paper P-3656, Institute for Defense Analyses, 44 pp.

Nuttall, A. H., 1966: On the quadrature approximation to the Hilbert transform of modulated signals. *Proc. IEEE*, **54**, 1458–1459.

Oppenheim, A., and R. Schafer, 1989: *Discrete-Time Signal Processing*. Prentice Hall, 879 pp.

Parks, T. W., and C. S. Burrus, 1987: *Digital Filter Design*. John Wiley & Sons, 342 pp.

Schwartz, M., 1990: *Information Transmission, Modulation, and Noise.* McGraw–Hill, 742 pp.

Reginald N. Meeson, Jr.
Institute for Defense Analyses, 4850 Mark Center Drive, Alexandria, VA 22311-1882, USA
meeson@ida.org