The Fast Fourier and Hilbert-Huang Transforms: A Comparison

Denis Donnelly

Abstract: The conversion of time domain data via the fast Fourier (FFT) and Hilbert-Huang (HHT) transforms is compared. The FFT treats amplitude vs. time information globally as it transforms the data to an amplitude vs. frequency description. The HHT is not constrained by the assumptions of stationarity and linearity, required for the FFT, and generates both amplitude and frequency information as a function of time. The behavior and flexibility of these two transforms are examined for a number of different time domain signal types.

Keywords: fast Fourier transform, Hilbert-Huang transform, data analysis

1   Introduction

A common approach in spectrum analysis for extracting frequency information from time series data is to use the fast Fourier transform (FFT) [DON, 05]. A more recent method [HUA, 98], [HUA, 05], which generates amplitude and frequency vs. time spectra, is the Hilbert-Huang transform (HHT). These two approaches are fundamentally different. The FFT assumes stationarity and linearity of the data and relies on globally defined orthogonal basis states. The HHT does not require the same assumptions of the data. In order to represent nonlinear and nonstationary data, global basis states must be replaced with adaptive, locally determined ones, a process the first stage of the HHT does perform. The resulting basis states are, in general, not strictly orthogonal.

After a very brief review of the FFT, the HHT is described. The comparative behavior of the two transforms is then explored. Various data sets with different signal characteristics are examined. Finally, the chirp of a bat is analyzed.

2   The Fast Fourier Transform

The fast Fourier transform (FFT) provides an efficient algorithm for converting data from the time domain into the frequency domain. Typically, the data to be transformed consists of N uniformly spaced points $x_j = x(t_j)$ where $N = 2^n$ with n an integer, and $t_j = j\Delta t$ where j ranges from 0 to $N - 1$. The discrete Fourier transform can be expressed in several ways. A commonly used form is the following (with $i = \sqrt{-1}$):

$$X_k = \sum_{j=0}^{N-1} x_j exp(-2\pi i \frac{j}{N} k) \tag{1}$$

where $k = -N/2, \ldots, -1, 0, 1, \ldots N/2 - 1$ and where x_j represent the time domain data and X_k their representation in the frequency domain. The algorithm for the FFT conversion process (Cooley-Tukey or any of several other comparable algorithms) makes the FFT widely applicable as it reduces the number of computations from something on the order of n^2 to $nlogn$ which obviously provides an enormous reduction in computation time.

The frequency data are typically displayed in one of two ways: an amplitude spectrum or a power spectrum. The amplitude spectrum is typically expressed by the relation

$$A_k = \frac{2}{N} |X_k|. \tag{2}$$

Whereas the power spectrum is typically expressed by the relation

$$P_k = \frac{1}{N} \left| X_k^2 \right|.$$

(3)

where $k = 0, 1, \ldots, N/2$.

3   The Hilbert-Huang Transform

The Hilbert-Huang transform is carried out in two stages: 1) the empirical mode decomposition (EMD) process, which deconstructs the signal into a set of intrinsic mode functions (IMF) and 2) the extraction of frequency vs. time information from each of the IMF's in combination with its Hilbert transform (HT). A brief summary of the process for discretely sampled signals follows.

The EMD process deconstructs the original signal into a set of IMF's. Each IMF, extracted from the signal by a series of siftings, has two fundamental properties: 1) the number of extrema and the number of zero crossings differ, at most, by one and 2) the mean value of the envelopes defined by the local maxima and local minima is zero. Unlike the harmonic functions of a Fourier series, these oscillatory functions may vary in both amplitude and frequency over time. In this decomposition process, the first IMF contains the highest frequencies associated with the original signal; each subsequent IMF contains lower-frequency components.

Amplitudes and frequencies are extracted from these IMF's in the second stage of the HHT process. The instantaneous amplitude and angular frequency associated with each IMF depend on the amplitude and phase of a complex number that the IMF and its Hilbert transform (HT) define. The real part of the complex number is the IMF; the imaginary part of the number is the IMF's HT. The instantaneous amplitude is the amplitude of this complex number. The instantaneous angular frequency associated with that IMF is the derivative of the unwrapped phase. The entire process is repeated for each IMF to extract the complete frequency versus time information from the original data.

The computation of the HT is essentially a convolution of an IMF, $x(t)$, with $1/t$. The effect of this convolving is to emphasize the local properties of $x(t)$. This locality preserves the time structure of the signal's amplitude and frequency.

The EMD process is designed to deconstruct the complete signal into a set of IMFs, each of which is extracted from a starting data set via a sifting process. This sifting process is repeated until the criteria listed above are satisfied and the difference between successive siftings is suitably small (there is some discussion as to what the terminating difference between siftings should be; it depends to some extent on the data set being examined). The process of extracting IMF's terminates when the residual contains no significant frequency information.

This sifting process comprises several steps. Given a discretely sampled signal $y(t)$,

1. Determine the location of all maxima, $ymax(t)$, and minima $ymin(t)$ $y(t)$.

2. Fit a cubic spline through the $ymax(t)$ and another through the $ymin(t)$.

3. Calculate the mean of the spline curves at each point

$$m(t) = (ymax(t) + ymin(t))/2$$

4. Remove the trend, $m(t)$. Let $d(t) = y(t) - m(t)$.

5. Is $d(t)$ an IMF? If $d(t)$ meets the criteria defining an IMF, let $c_i(t) = d(t)$ and advance i by 1. Extract the residual $r(t) = y(t) - d(t)$. If $d(t)$ doesn't meet the criteria, further sifting is required. Repeat steps 1 through 5, substituting $d(t)$ for $y(t)$.

6. Repeat steps 1 through 5 until the residual no longer contains any useful frequency information.

The original signal is, of course, equal to the sum of its parts. If we have N IMFs and a final residual $r_N(t)$,

$$y(t) = \sum_{i=1}^{N} c_i(t) + r_N(t) \tag{4}$$

The second stage of the HHT process extracts the amplitude and frequency information from each IMF. The steps for amplitude and frequency extraction from a given discrete IMF are as follows:

1. Compute the IMF's discrete Fourier transform (DFT) using the series expression (1) for the transform.

2. Compute the HT. Use the real and imaginary parts of step 1's DFT as coefficients ($M = N/2$):

$$y_j = \frac{1}{N} \sum_{k1=0}^{M} \left(Re(X_{k1} \sin(2\pi k1 \frac{j}{N}) + Im(X_{k1} \cos(2\pi k1 \frac{j}{N}) \right)$$
$$+ \left(\frac{-1}{N} \right) \sum_{k2=M+1}^{N-1} \left(Re(X_{k2} \sin(2\pi k2 \frac{j}{N}) + Im(X_{k2} \cos(2\pi k2 \frac{j}{N}) \right) \tag{5}$$

3. Form the complex number $z_j = x_j + iy_j$, extract the phase $\phi_j = tan^{-1}(y_j/x_j)$.

4. Unwrap the phase so that it becomes a monotonically increasing function.

5. Determine the frequency. Take the derivative of the phase

$$f_j = \frac{1}{2\pi} \frac{d\phi_j}{dt}$$

6. Determine the amplitude.

$$a_j = \sqrt{x_j^2 + y_j^2}$$

4   Examples

The first examples treat cases where the signal in the time domain has the properties of an IMF and no sifting is required. The examples are chosen to demonstrate the responses of the FFT and HHT to changes in frequency and amplitude of the time domain signal.

The first signal considered consists of two cycles of a sine wave with frequency equal to one followed after a brief pause by four cycles of a sine wave with a frequency equal to two. Fig. 1 shows both the signal and its HHT. The oscillations in the HHT are typical and vary only slightly with the amplitude of the original signal. The low frequency end of the FFT for this example is shown in Fig. 2. While it does show the major peaks at frequencies equal to one and two, it has a rather rough appearance. The complex structure of the FFT arises because, in effect, the FFT has to create a Fourier series to fit the signal comprising both the two frequencies as well as the portions where the signal is zero.

The second signal has a fixed frequency sine wave with amplitude increasing with time. The HHT, Fig. 3, is not daunted by this amplitude variation. The FFT on the other hand has an amplitude which is consistent with an average value of the peak amplitudes for the time domain data.

The third signal, a chirp, has a constant amplitude and a variable frequency which increases over time by one order of magnitude. The HHT provides an accurate representation of the frequency as long as the signal is sampled with sufficient frequency. If the number is too small the oscillations in the frequency curve grow and can have negative values. Fig. 4 shows the frequency response; here N=2048 and there

are more than twenty computed sample points in the last (highest frequency) cycle. The FFT (Fig. 5.) is nearly constant for the frequency spanned but drops off near the ends.

An example which clearly contrasts the behavior of the FFT and HHT and which demonstrates which process is more correct physically, is that of a sum of two sine waves with frequencies close enough to readily show a beat pattern (f1 = 1, f2 = 1.08). The FFT of such a pattern is shown in Fig. 6. The two original frequencies are clearly present in the amplitude spectrum. The HHT on the other hand generates a single frequency of 1.04 which is an average of the two sine wave frequencies. (When dealing with signals that show a beat pattern, for example, artifacts are likely to appear in the frequency representation near the nodal points of the signal.) The HHT amplitude function follows the beat pattern as an envelope. While the description of the beat signal as a sum of two individual frequency signals is mathematically equivalent to a single signal with a frequency equal to the average of the two frequencies used to construct the signal, times an amplitude modulating function with a frequency equal to the difference in frequency of the two signals, the two different representations are not equivalent physically. The HHT provides the physically correct frequency description; the instantaneous frequency is the average frequency. This is an important notion that is worth emphasizing. We are used to talking about the harmonic content of signals, which is a mathematically correct description, but it may not be the best representation of what is actually taking place.

It is interesting to observe what happens to the amplitude spectrum of the FFT, as a smaller and smaller number of cycles of the pattern are subjected to analysis. If the number of cycles is small enough that the signal does not go beyond the first node of the envelope, then the FFT shows only one peak at the average value of the two frequencies.

The HHT cannot treat functions such as a single cycle of a square wave as there are no maxima and minima to fit, while the FFT can readily analyze such a signal. If a small amplitude sine wave is made to ride on top of a square wave, for example, the HHT will recognize the sine wave away from the transition region. Near the step transition, the frequency response shows a large spike.

It is useful to know how differing frequencies are distributed during the EMD process. An estimate can be obtained by taking the FFT's of the first and second IMF's as the difference between two frequencies is increased. From this, we can observe the behavior of the sorting process. For example, a factor of two or greater in the frequencies is needed to shift the lower frequency signal well into the next IMF (see Table 1). For example, if the EMD process is applied to a signal comprising frequencies 1 and 0.5, the first IMF yields an average frequency value of 1 with fluctuations ranging approximately from 0.96 to 1.05; the second IMF yields an average value of 0.5 with the fluctuations ranging from 0.49 to 0.51.

Frequency 2	Peak Amplitude Frequency 1 = 1	Peak Amplictude Frequency 2 IMF1	Peak Amplitude Frequency 2 IMF2
0.95	526	489	
0.9	521	482	
0.8	516	444	14
0.7	513	362	55
0.6	510	188	178
0,5	507	34	327
0.4	508		409

For a final application, we look at the chirp of a bat: 400 data points sampled at 7 intervals. The bat's chirp does consist of frequencies decreasing over time, consequently, an FFT of the entire signal would not yield useful information. Instead, to use the FFT, a sliding window approach is used [DON, 06]. For example, look at 64 point subsets of the entire data set. In this time range, the inherent frequency structure is revealed. As this short window slides along, the changing frequencies can be tracked. The HHT of the chirp, shows information of interest in the first and second IMF's. Results are shown in Fig. 7. The

second harmonic had too small an amplitude for the HHT process to recognize it. Both analysis forms would benefit from a shorter sampling interval. In the figure, the fit to the fundamental is proportional to $exp(-const\sqrt{t})$. The fit to the first harmonic is simply two times that of the fundamental. Because of the complexity of the signals in the figure, the FFT results for the sliding window technique are not shown. But they approximate the fitted lines shown in the figure. There is a weak second harmonic which can be extracted from the data when using the FFT process. However, that signal is too weak for the HHT to recognize.

5   Final Comments

The combination of sifting associated with the EMD process, and then taking the HT, unwrapping the phase, and taking the derivative of each IMF is computationally much more demanding than taking the FFT. However, the HHT not only provides instantaneous frequency and amplitude descriptions, it results in representations that are more meaningful physically.
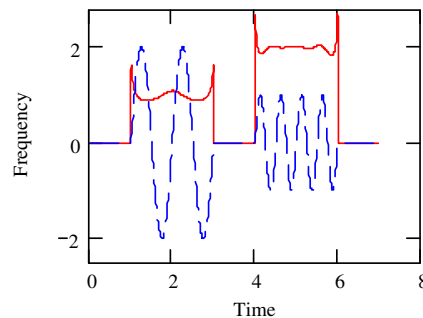
Figure 1: HHT frequency representation (solid line) of the time domain signal (dashed).
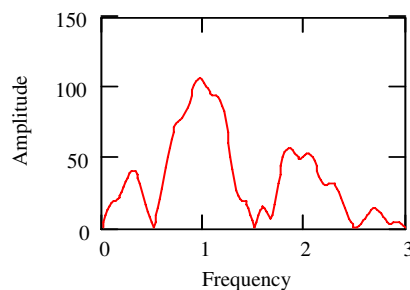

Figure 2: FFT amplitude spectrum (solid line) of the (dotted) signal shown in Fig. 1. Only the lower frequencies are displayed for emphasis.
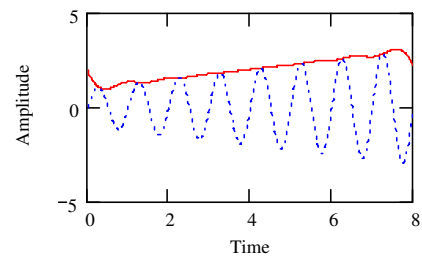
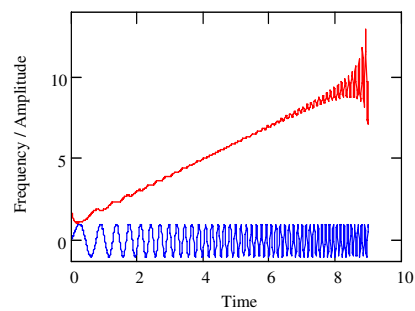Figure 3: HHT amplitude spectrum (solid line) of the (dotted) signal.

Figure 4: HHT frequency representation of a chirp signal. The HHT follows the increase in chirp frequency as it scales from one to ten. The amplitude one chirp signal is also shown.
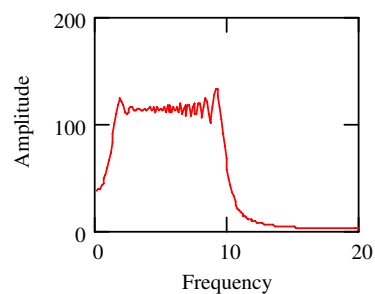

Figure 5: FFT of a chirp signal with a frequency ranging from one to ten. Only the lower frequencies are displayed for emphasis.
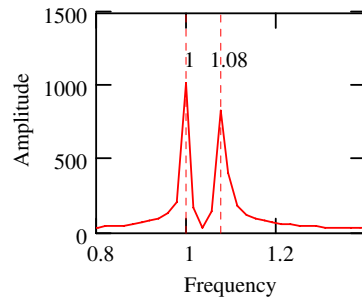
Figure 6: The FFT of a signal consisting of the sum of two sine waves with frequencies 1 and 1.08. Only a portion of the FFT is shown for emphasis.
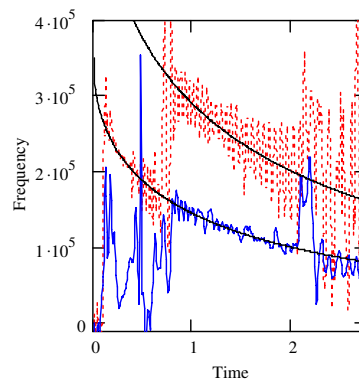

Figure 7: Frequency vs. time of a bat chirp as determined via HHT. The fluctuating curves represent the HHT results. The solid curve corresponds to the fundamental. The dotted curve is the first harmonic. The two smooth curves are fits to the data.

References

[1] Denis Donnelly and Bert Rust, "The Fast Fourier Transform for Experimentalists: Part I Concepts," Computing in Science & Eng., vol. 7, no. 2, pp. 80-88, 2005.

[2] Denis Donnelly, "The Fast Fourier Transform for Experimentalists: Part VI Chirp of a Bat," Computing in Science & Eng., vol. 8, no. 2, pp. 72-78, 2006.

[3] Norden E. Huang et. al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. Lond. A., vol. 454, pp. 903-995, 1998.

[4] Norden E. Huang and Samuel Shen, Eds. New Jersey: World Scientific, 2005.

Denis Donnelly
Siena College
Department of Physics
Loudonville, NY 12211
E-mail: donnelly@siena.edu