A generalized demodulation approach to time-frequency projections for multicomponent signals

BY S. OLHEDE AND A. T. WALDEN
Department of Mathematics, Imperial College London,
180 Queen’s Gate, London SW7 2BZ, UK
(a.walden@ic.ac.uk)

In this paper, we introduce a flexible approach for the time-frequency analysis of multicomponent signals involving the use of analytic vectors and demodulation. The demodulated analytic signal is projected onto the time-frequency plane so that, as closely as possible, each component contributes exclusively to a different ‘tile’ in a wavelet packet tiling of the time-frequency plane, and at each time instant, the contribution to each tile definitely comes from no more than one component. A single reverse demodulation is then applied to all projected components. The resulting instantaneous frequency of each component in each tile is not constrained to a set polynomial form in time, and is readily calculated, as is the corresponding Hilbert energy spectrum. Two examples illustrate the method. In order better to understand the effect of additive noise, the approximate variance of the estimated instantaneous frequency in any tile has been formulated by starting with pure noise and studying its evolving covariance structure through each step of the algorithm. The validity and practical utility of the resulting expression for the variance of the estimated instantaneous frequency is demonstrated via a simulation experiment.

Keywords: demodulation; Hilbert spectrum; instantaneous frequency; multicomponent signals; wavelet packets; wavelet transform

1. Introduction

The analysis of signals with time-evolving oscillatory structure is an area of active interest. Basic representations of such signals are bilinear distributions (e.g. Cohen 1995), which are calculated by filtering the signal and its complex conjugate with a kernel function. This generally entails problems for signals with multiple components, unless the kernel decays very rapidly in time-frequency, thus avoiding damaging interference effects. Our aim in this work is to introduce a simple approach to the problem of tracking the time-dependent frequency content of each component of a multicomponent signal via the undecimated discrete wavelet packet transform. It is well known that wavelet transforms ‘tile’ the time-frequency plane in a proportional bandwidth or octave band manner. The extension to wavelet packets allows more irregular, but still rectangular, partitioning of the time-frequency plane. Such rectangular tilings of the
time-frequency plane are not best suited to representing some classes of signals (such as FM signals), which may have energy distributions which are slanted, nonlinear, etc., in the time-frequency plane.

Non-rectangular tilings can be designed. Baraniuk & Jones (1993) designed scale-shear fan bases corresponding to tiling with truncated wedges, well matched to representing linear chirp signals with few expansion coefficients. Chirplet transforms (Mann & Haykin 1992, 1995; Baraniuk & Jones 1995, 1996) provide a systematic framework for designing new signal representations, and can be extended to deal with not just slant but also curvature in the time-frequency plane. None the less, a general approach to discrete chirplet transforms is not forthcoming. Gaussian chirplets do not form an orthogonal basis, and adaptive matching pursuit Gaussian chirplet decompositions have thus been advocated (Yin et al. 2002). Philosophically related is research by Hlawatsch et al. (1999), Papandreou-Suppappola et al. (2001) and Papandreou-Suppappola & Suppappola (2002), which explore time and frequency warping.

The fractional Fourier transform (FrFT), probably more widely used in optical research, is also closely related to wavelet and chirp transforms (Ozaktas et al. 1994). The Fourier transform of fractional order $a$ is defined in such a way that the standard Fourier transform is a special case with order $a=1$, while successively transforming with fractional orders $a_1$ and $a_2$ is the same as transforming with order $a_1 + a_2$ (so that the standard Fourier transform corresponds to two successive half-transforms). Ozaktas et al. (1994) pointed out that the effect of the FrFT is to rotate the Wigner distribution of a signal, and showed, that consequently, linear chirp functions are simply the time domain representation of signals that appear as delta functions or harmonics in other fractional domains. Similarly, Capus et al. (2000) noted that the FrFT is composed of a multiplication by a linear chirp in the time domain, followed by Fourier transformation, then by multiplication by a linear chirp in the transformed domain, and, finally, by a complex scaling. Mendlovic et al. (1997), see also Huang & Suter (1996), suggested a ‘fractional wavelet transform’ algorithm, which consists of an FrFT of order $a$ followed by continuous wavelet transformation; for reconstruction, use the inverse wavelet transform, followed by an FrFT of order $-a$. If the signal was dominated by a linear chirp, then the value of $a$ should match the slant in the time-frequency plane. In general, the order $a$ is chosen by optimization, and it is suggested that the high computational burden can be reduced by use of optical processing.

Other recent work includes that by Stevenson et al. (2001) and Capus & Brown (2003). Our approach can be viewed as a Fourier/wavelet packet transform hybrid. Instead of using undecimated wavelet packet coefficient sequences, it uses undecimated wavelet packet detail sequences, each of which can be viewed as arising by Fourier transforming the signal, multiplying it by the modulus squared of the Fourier transform of a band-pass filter, followed by inverse Fourier transformation. Our approach to generalizing this approach is to replace the standard Fourier transformation by the generalized Fourier transformation (GFT) of Detka & El-Jaroudi (1996). For a signal $x(t)$, the GFT is given by $X_G(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t + s_0(t)} dt$, where $s_0(t)$ is a real-valued function depending on time only, and specifies the evolutionary phase behaviour of the signal. Note this is the same as applying the standard Fourier transform to $x(t) e^{-i2\pi s_0(t)}$. With the inverse GFT, we apply the usual inverse Fourier...
transform to $X_G(f)$, followed by multiplication of the result by $e^{i2\pi s_0(t)}$ because

$$x(t) = \int_{-\infty}^{\infty} X_G(f)e^{i2\pi f + s_0(t)} df = e^{i2\pi s_0(t)} \int_{-\infty}^{\infty} X_G(f)e^{i2\pi f} df.$$ (1.1)

Hence, if $X_G(f) \equiv \delta(f-f_0)$, then $x(t) = e^{i2\pi (f_0 + s'_0(t))}$, that is, a signal with instantaneous frequency $v(t) = f_0 + s'_0(t)$ will be mapped to the point $f = f_0$, where $s'_0(t) = ds_0(t)/dt$. Therefore, if we wish to map a signal with, for example, a curved path in the time-frequency plane specified by $f_0 + s'_0(t)$ into a wavelet packet passband (rather than a single point), we simply need to specify a function $s(t)$ approximating $s_0(t)$ such that the demodulated function $x(t)e^{-i2\pi s(t)}$ will be mapped into the passband. Of course, this result is very flexible in that $s'_0(t)$ is not constrained to be linear or even quadratic.

Many audio-, acoustic- and speech-type signals can be characterized as multicomponent signals, with time-frequency distribution consisting of several parallel (or similarly oriented) linear or curved energy paths with low instantaneous bandwidth. Examples abound: Flandrin (1988) examined bat sonar, Cohen (1989, 1992) considers seismic signals and whale sounds, Sucic & Boashash (2002) look at bird song, and Gribonval & Bacry (2003) consider musical recordings. Our approach is particularly well suited to (and most easily implemented for) such signals because with the demodulation, the components should generally be shiftable into different wavelet packet frequency bands.

In order to track the time-dependent frequency content of each component of a multicomponent signal, we shall first demodulate the complete signal, then project it onto the time-frequency plane in a manner such that (i) as closely as possible, each component contributes exclusively to a different ‘tile’ in the chosen (wavelet packet) tiling of the time-frequency plane, and (ii) at each time instant, the contribution to each tile definitely comes from no more than one component. The projected components are ‘analytic’ vectors. The single reverse demodulation is then applied to all projected components. The instantaneous frequency of each component in each tile is then well defined, and can be calculated and weighted by the energy to yield the ‘Hilbert energy spectrum’ for that projection. Agglomeration over projections yields the complete Hilbert spectrum (Huang et al. 1998; Olhede & Walden 2004a).

2. The generalized demodulation algorithm

We give details of our algorithm using the more practically useful discrete-time notation. We assume we have sampled a continuous-time (multicomponent) signal $x(t)$ with a sample interval of unity to get an even-length vector of observations $\mathbf{x} = [x_0, \ldots, x_{N-1}]^T$ and have done likewise to $s(t)$ to get $\mathbf{s} = [s_0, \ldots, s_{N-1}]^T$. For a sample interval of unity, the Nyquist frequency is $f = 1/2$.

(a) Analytic vectors and forward demodulation

We firstly create an ‘analytic’ vector $\mathbf{y} = \mathbf{x} + i\mathcal{H}\mathbf{x}$, where $\mathcal{H}$ is the discrete Hilbert transform (DHT) matrix (Marple 1999), and then apply the forward demodulation to obtain $d_l = y_le^{-i2\pi s_l}$, $l = 0, \ldots, N-1$. We can write this in matrix
form as $\mathbf{d} = (C - i\delta)\mathbf{y}$, where $(C)_{lm} = \cos(2\pi s_l)\delta_{lm}$, and $(\delta)_{lm} = \sin(2\pi s_l)\delta_{lm}$, where $\delta_{lm}$ is the Kronecker delta, and $m = 0, \ldots, N - 1$.

The discrete Fourier transform of an analytic vector has zero coefficients for negative frequencies. Hence, by working with the analytic vector, any energy in $\mathbf{x}$ in the negative frequencies is not inadvertently moved into the positive frequencies by the demodulation. The new complex-valued vector $\mathbf{d}$ is not analytic, and so we now create a new analytic vector $\mathbf{z} = \mathbf{d} + i\mathcal{H}\mathbf{d}$.

We assume $\mathbf{x}$ to be band-limited to $\left[-\frac{1}{2} + \frac{1}{N}, \frac{1}{N} \right] \cup \left[\frac{1}{N}, \frac{1}{2} - \frac{1}{N} \right]$ for discrete computation of the analytic vector $\mathbf{y}$. The effect of the demodulation is to change frequencies by $-s'_l$, where $s'_l$ denotes the discrete version of $s'(t)$. When the new analytic vector $\mathbf{z}$ is created, we must assume that $\mathbf{d}$ is band-limited to $\left[\frac{1}{N}, \frac{1}{2} - \frac{1}{N} \right]$, or, equivalently, that $\mathbf{y}$ is band-limited to $\left[\frac{1}{N} + s'_l, \frac{1}{2} + s'_l \right]$, for $l = 0, \ldots, N - 1$. This is the same as saying that the demodulation must not move signal component energy out of the frequency range $\left[\frac{1}{N}, \frac{1}{2} - \frac{1}{N} \right]$ and this is simple to check graphically, as shown by the examples of §3.

(b) Wavelet packet projections

The particular wavelet-type transform we will apply to $\mathbf{z}$ is the so-called maximal overlap discrete wavelet packet transform (MODWPT; Walden & Contreras Cristán 1998). This is an undecimated transform. Each MODWPT is associated with a transform level $j$, ($j = 1, \ldots, J_0$), and the $j$th level decomposes the frequency interval $[0, 1/2]$ into $2^j$ equal passband intervals $g_{j,n} = \left(\frac{n}{2^j \pi}, \frac{n + 1}{2^j \pi}\right)$, where $n$, ($n = 0, \ldots, 2^j - 1$), denotes the frequency band index within level $j$. For any $j$, the $g_{j,n} \times [0, (N - 1)]$ partitions (tile) the time-frequency plane into equal width rectangles. Of course, a larger value of $j$ corresponds to a narrower passband, and our interest will be in the final level $J_0$ MODWPT, which decomposes the frequency interval $[0, 1/2]$ into $2^{J_0}$ equal passband intervals; the choice of $J_0$ is considered in §2f.

For the MODWPT coefficients indexed by $(j, n)$, we write $\tilde{\mathbf{W}}_{j,n} = \tilde{\mathbf{W}}_{j,n}\mathbf{z}$, where $\tilde{\mathbf{W}}_{j,n}$ is an $N$-length column vector, and $\tilde{\mathbf{W}}_{j,n}$ is an $N \times N$ transformation matrix. The wavelet packet coefficients for $t = 0, \ldots, N - 1$ can be written as $\tilde{\mathbf{W}}_{j,n,t} = \sum_{l=0}^{L_j - 1} \tilde{u}_{j,n,l}z(t - l)_{\text{mod } N}$, where $\{\tilde{u}_{j,n,l}\}$ is a $j$th level and $n$th band, MODWPT wavelet packet filter of length $L_j = (2^j - 1)(L - 1) + 1$, derived from the basic DWT scaling and wavelet filters $\{g_l, l = 0, \ldots, L - 1\}$ and $\{h_l, l = 0, \ldots, L - 1\}$, as described in Percival & Walden (2000, §6.6). The tilde notation is used to distinguish the MODWPT filters and coefficients from the decimated (DWT) equivalents, as in Percival & Walden (2000).

Let $\tilde{u}_{j,n,l}^*$ be the filter obtained by periodizing $\tilde{u}_{j,n,l}$ to length $N$; that is, $\tilde{u}_{j,n,l}^* = \sum_{m=0}^{\infty} \tilde{u}_{j,n,l+mN}$. Then, an equivalent representation is

$$\tilde{\mathbf{W}}_{j,n,t} = \sum_{l=0}^{N - 1} \tilde{u}_{j,n,l}^*z(t - l)_{\text{mod } N}. \quad (2.1)$$

We will actually make use of sequences $\tilde{\mathbf{D}}_{j,n} = \tilde{\mathbf{W}}_{j,n}^T, \tilde{\mathbf{W}}_{j,n}\mathbf{z} = \tilde{\mathbf{W}}_{j,n}^T \tilde{\mathbf{W}}_{j,n}$ with $j = J_0, n = 1, \ldots, 2^{J_0} - 1$. For $n > 0$, the sequence $\{\tilde{\mathbf{D}}_{j,n,l}, l = 0, \ldots, N - 1\}$ consists of level $j$,
frequency band \( n \), MODWPT detail coefficients. (The sequence \( \{\tilde{D}_{j,0,l}, l=0, \ldots, N-1\} \) involves only low-pass filters and covers low frequencies and trends.)

For a level \( j \), the analytic vector can be recovered via \( z = \sum_{n=0}^{2^j-1} \tilde{D}_{j,n} \) so that if we define projection matrices \( P_{j,n} = {\tilde{W}}_{j,n}^T (Olhede & Walden 2004) \), then \( z = \sum_{n=0}^{2^j-1} P_{j,n} z = \sum_{n=0}^{2^j-1} \tilde{D}_{j,n} \); that is, we can decompose \( z \) into a linear combination of the contributions to \( z \) in projection subspaces that tile the time-frequency plane via a set of equal bandwidth rectangles. We note that at any time \( l, l=0, \ldots, N-1 \), energy in \( x \) in the frequency band \( \left( \frac{m}{2^j}, \frac{m+1}{2^j} \right) \), is mapped to \( j_{j,n} \).

Any MODWPT detail sequence \( \tilde{D}_{j,n} \), created as above, is an analytic vector. To see this, note that because \( z = d + i \mathcal{H} d \), we find that \( \tilde{D}_{j,n} = P_{j,n} z = P_{j,n} d + i P_{j,n} \mathcal{H} d \). However, the matrices \( P_{j,n} = {\tilde{W}}_{j,n}^T {\tilde{W}}_{j,n} \) and \( \mathcal{H} \) commute (see appendix A) so that \( \tilde{D}_{j,n} = P_{j,n} d + i \mathcal{H} P_{j,n} d \) and hence \( \tilde{D}_{j,n} \) is an analytic vector.

(c) **Reversed demodulation**

Any details vector \( \tilde{D}_{j,n} \) produced as above can be viewed as having been created by (discrete) GFT, followed by multiplication with the modulus squared of the Fourier transform of a band-pass filter, and then by inverse Fourier transformation. To turn this last step of inverse Fourier transformation into inverse GFT, we carry out a reverse demodulation (see equation (1.1)) via \( \mathcal{Y}_{j,n,l} = \tilde{D}_{j,n,l} e^{i 2 \pi s_l}, l=0, \ldots, N-1 \).

(d) **The Hilbert spectrum and instantaneous frequency**

For any projection indexed by \( (j, n) \), the Hilbert spectrum (Huang et al. 1998; Olhede & Walden 2004) can be formed by firstly calculating the amplitude and phase sequences given by

\[
A_{j,n,l} = \sqrt{\Re^2 \{\mathcal{Y}_{j,n,l}\} + \Im^2 \{\mathcal{Y}_{j,n,l}\}} \quad \text{and} \quad \phi_{j,n,l} = \tan^{-1} \left[ \frac{\Im \{\mathcal{Y}_{j,n,l}\}}{\Re \{\mathcal{Y}_{j,n,l}\}} \right].
\]

The derivative of the phase at discrete time \( l \) may be calculated via a fourth-order generalized phase-difference estimator (Boashash 1992, p. 542) so that the instantaneous frequency is given by

\[
u_{j,n,l} = \frac{1}{2\pi} \sum_{m=-2}^{2} r_m \phi_{j,n,l-m} \]

with \( r_{-1} = \pm (8/12), r_{1,2} = \pm (1/12) \).

Now, for a large \( M \), define \( M \) frequencies \( f_k = k \Delta f, \ k = 0, \ldots, M-1 \), where \( \Delta f = 1/[2(M-1)] \), covering the interval \([0,1/2]\). Then, the Hilbert energy spectrum of \( \{\mathcal{Y}_{j,n,l}\} \) may be defined as

\[
S_{j,n,l,f_k} = A_{j,n,l}^2 \delta_k (r_{j,n,l}/\Delta f), \quad l=0, \ldots, N-1; \ f_k = 0, \ldots, 1/2.
\]

Here, \( \delta_{k,m} \) is the Kronecker delta, \( \langle x \rangle \) denotes the integer closest to \( x \); we used \( M=512 \).
The algorithm summarized

(i) Create the analytic vector \( y = x + i \mathcal{H} x \);
(ii) Forward demodulate using \( d = (C - i S) y \);
(iii) Create new analytic vector \( z = d + i \mathcal{H} d \);
(iv) Compute MODWPT projections \( \tilde{D}_{j_0,n} = P_{j_0,n} z \), \( n = 1, \ldots, 2^{J_0} - 1 \);
(v) Carry out reverse demodulation: \( Y_{J_0,n} = (C + i S) \tilde{D}_{j_0,n} \);
(vi) Compute the instantaneous frequency sequences \( \{ \nu_{j_0,n,l} \} \), \( n = 1, \ldots, 2^{J_0} - 1 \) according to equation (2.3). Compute and plot the Hilbert energy spectrum as detailed in equation (2.4).

The choice of \( J_0 \)

In our algorithm, we will compute MODWPT projections for level \( j = J_0 \), and so we need to choose \( J_0 \). Using, as in Olhede & Walden (2004a), the optimal asymptotic frequency resolution filters \( \{ g_l \}, l = 0, \ldots, L - 1 \), known as Fejér-Korovkin filters, the band-pass concentration of the MODWPT filters \( \tilde{u}_{j_0,n,l} \) is very good, having the property that the only significant leakage from band \( n \) is into band \( n - 1 \) and/or \( n + 1 \), but no further. This is true at least for \( J_0 = 2, 3 \) with \( L = 18 \) and 4 with \( L = 22 \). (See fig. 4 of Olhede & Walden (2004a) for \( J_0 = 3 \).) \( J_0 = 2, 3 \) and 4 corresponds to dividing \([0,1/2] \) into 4, 8 or 16 frequency bands. Our ‘rule’ is to choose \( J_0 \) as small as possible, consistent with having at least one empty projection band between bands containing signal from different components in step 4 of the algorithm of §2e. In the examples that follow, it is seen that this strategy can be achieved by taking \( J_0 = 3 \). Since the leakage properties hold at least to the 16 band case of \( J_0 = 4 \), the minimum frequency separation between components suitable for our algorithm is one-sixteenth of the Nyquist frequency.

3. Examples

(a) Application to real data

Our first example is that of the echolocation pulse emitted by the large brown bat, \textit{Eptesicus fuscus}, digitized with a sample interval of \( \Delta t = 7 \) \( \mu \)s, with \( N = 400 \). Figure 1a shows the standard Wigner-Ville distribution for the analytic signal vector \( y \). The \( p = 3 \) significant components generate \( p(p - 1)/2 = 3 \) ‘outer’ interferences (i.e. interferences due to interaction of components; Flandrin 1999, p. 232), and there is also visible ‘inner’ interference where the time-frequency energy distributions are noticeably nonlinear. Figure 1b shows the ordinary Hilbert energy spectrum via MODWPT projections, \textit{excluding any demodulation}; that is, steps 2, 3 and 5 are omitted in the algorithm of §2e. Here, \( j = J_0 = 3 \) so that \( g_{j,n} = \left( \frac{n}{2^{j+1} \Delta t}, \frac{n+1}{2^{j+1} \Delta t} \right) = \left( \frac{n}{16 \Delta t}, \frac{n+1}{16 \Delta t} \right) \), for \( n = 0, \ldots, 7 \). These eight frequency passbands are delineated by the horizontal dotted lines; the band between zero and the first horizontal dotted line corresponds to \( n = 0 \), with increasing \( n \) until the top interval between the uppermost horizontal line and the top of the plot which covers the band \( n = 7 \). Although, as required in multicomponent analysis, \textit{at each time instant}, the different frequency chirps have been successfully separated into different frequency sub-bands; each component spans at least two
sub-bands and is noticeably affected by the band-edge imperfections in the (albeit extremely good) band-pass filters generated using Fejér–Korovkin filter coefficients (see Olhede & Walden 2004a).

Turning now to demodulation, the form of $s_l$ used here was $s_l = b_2 t^2 + b_1 l$, $l = 0, \ldots, N-1$, where $b_2 = -0.0003$ and $b_1 = 0.175$; hence, $s'_l = (2b_2 l + b_1) / \Delta t$. The value of $b_2$ is easily ‘guessed’ from the Hilbert spectrum results of figure 1b by following the middle component, and $b_1$ can then be chosen suitably to align the signal components within the passbands as shown via the diagonal dashed lines in figure 1b. The bottom dashed line shows $\frac{n}{2^{n+1} \Delta t} + s'_l$, $l = 0, \ldots, N-1$ for $n=1$; that is, $\frac{1}{16 \Delta t} + s'_l$, $l = 0, \ldots, N-1$. The next to bottom line shows $\frac{n+1}{2^{n+1} \Delta t} + s'_l$, $l = 0, \ldots, N-1$, for $n=1$, i.e. $\frac{2}{16 \Delta t} + s'_l$, $l = 0, \ldots, N-1$. Thus, the interval delineated by the two bottom lines is $\left( \frac{n}{2^{n+1} \Delta t} + s'_l, \frac{n+1}{2^{n+1} \Delta t} + s'_l \right)$, $l = 0, \ldots, N-1$, for $n=1$ and $j = J_0 = 3$, and as pointed out above, energy in

**Figure 1.** (a) Wigner-Ville distribution corresponding to analytic version $y$ of the bat echolocation pulse; (b) the Hilbert spectrum using the MODWPT, 18-point Fejér–Korovkin wavelet filters, and $J_0=3$; (c) the Hilbert spectrum incorporating only forward demodulation; (d) the Hilbert spectrum using the complete algorithm. Details of marked lines, etc., are given in the text.
such an interval will be mapped to \( g_{j,n} = g_{3,1} = \left( \frac{1}{10\Delta t}, \frac{2}{10\Delta t} \right) \) following the demodulation in step 2 of the algorithm of §2e. This interval is marked with ‘\( n = 1 \)’ on the right of the plot. The corresponding intervals for \( n = 3 \) and \( n = 5 \) are also marked on the right of figure 1b and, following the demodulation, energy in these intervals will be mapped to \( g_{3,3} \) and \( g_{3,5} \).

Figure 1c shows the Hilbert spectrum via MODWPT projections incorporating forward demodulation only; that is, only step 5 is omitted from the algorithm in §2e. We note as predicted from figure 1b, the energy in the signal has been mapped to the passbands \( g_{3,1}, g_{3,3} \) and \( g_{3,5} \). The demodulation has succeeded in realigning the time-frequency energy to lie (almost perfectly) within the frequency sub-bands. Figure 1d shows the extremely good results obtained using the full algorithm incorporating both the forward and reverse demodulation steps. The curvature of the lower frequency component is well preserved—there is no assumption of linear chirp components.

Our chosen demodulation has kept signal component energy in the frequency range \( \left[ \frac{1}{7\Delta t}, \frac{1}{23\Delta t} - \frac{1}{7\Delta t} \right] \), as required (see §2a). We note that a poor choice of \( s_l \) could cause energy loss by not satisfying this requirement, or result in one or more demodulated components crossing band boundaries and suffering band edge-effects as seen in the basic Hilbert spectrum without demodulation in figure 1b.

\[ (b) \text{ Application to synthetic data} \]

Here, we consider a synthetic signal that closely resembles a whale vocalization in shape. It consists of two quadratic chirps, one having a shorter duration than the other:

\[
x_l = \begin{cases} \sin(a_1 l^3 + a_2 l^2 + a_3 l), & l = 0, \ldots, 249 \\ \sin(a_1 l^3 + a_2 l^2 + a_3 l) + \sin(a_4 (l - 512)^3 + a_5 l), & l = 250, \ldots, 1023, \end{cases}
\]

with \( a_1 = -1.256 \times 10^{-6}, \ a_2 = 0.001 \, 9, \ a_3 = 1.198 \, 3, \ a_4 = -1.278 \times 10^{-6}, \ a_5 = 1.382 \, 3 \). Figure 2a shows the standard Wigner–Ville distribution for the analytic signal vector \( y \). The two components generate a single case of outer interference (between the components), while the notable curvature of the components generates severe inner interference. Figure 2b shows the Hilbert spectrum via MODWPT projections, excluding any demodulation. Here, \( j = J_0 = 3 \) again so that, with \( \Delta t = 1, \ g_{j,n} = \left( \frac{n}{16}, \frac{n+1}{16} \right) \), for \( n = 0, \ldots, 7 \); these 8 frequency passbands are delineated by the horizontal dotted lines. Although, as with the bat signal, at any time the different frequency chirps have been successfully separated into different frequency sub-bands, here both components cross three sub-bands leading to band-edge imperfections. To eliminate these artefacts, we can apply the demodulation method. To find \( \{ s_l \} \) we defined the quadratic \( s_l' = 3b_3 l^2 + 2b_2 l + c \) and found the parameter values \( b_3, b_2, \ c \) by matching of its values at \( l = 0, 500 \) and 1000 to \( f = 0.125, \ 0.281 \) 25 and 0.14, read off the plot by eye; this choice of points corresponds roughly to an arc halfway between the two components. Then, we let \( s_l = b_3 l^2 + b_2 l + b_1 \) so that \( s_l' = 3b_3 l^2 + 2b_2 l + b_1 \), where \( b_1 \) is chosen to align the signal components sensibly.
within the passbands, as shown via the curved dashed lines in figure 2b. The values used were \( b_1 = -8.561 \times 10^{-2}, \) \( b_2 = 3.059 \times 10^{-4} \) and \( b_3 = -1.987 \times 10^{-7}. \) We have not used the known simulation parameters \( a_1, \ldots, a_5 \) to obtain \( s_h \) but worked entirely from the Hilbert spectrum of figure 2b, as would be the case in practice.

Figure 2c shows the Hilbert spectrum via MODWPT projections incorporating forward demodulation only. The energy in the signal has been mapped to the passbands \( J_{3,2} \) and \( J_{3,4}. \) The time–frequency energy lies comfortably within the frequency sub-bands. Also shown by the dash–dotted lines in figure 2c are the instantaneous frequencies calculated directly from the two separate components of the synthetic whale vocalization signal.

Our chosen demodulation has kept signal component energy in the frequency range \( \left[ \frac{1}{N}, \frac{1}{2} - \frac{1}{N} \right], \) as required.

Figure 2d shows the final result using both forward and reverse demodulation steps. Most of the artefacts visible in figure 2b have been completely eliminated,
and the result is excellent. Moreover, the instantaneous frequencies are also plotted as dash–dotted lines in this figure, but cannot be distinguished from the Hilbert spectrum calculated with our algorithm.

4. Standard deviation of the instantaneous frequency estimate in the presence of noise

In §2e, we gave the steps in the demodulation algorithm, ending with the calculation of instantaneous frequency (and thus the Hilbert spectrum). Hypothetically, some bias could be introduced in the calculation of the instantaneous frequency of a component, through the steps of the deterministic algorithm of §2e. However, the whale vocalization was shown to exhibit no significant bias, and so it is not inevitable. Minimization of wavelet packet filter leakage effects by choosing $J_0$ as described in §2f, and choosing $s_l$ to align the signal components comfortably within (i.e. not on the edge of) the passbands will always act to reduce bias.

We will now derive an expression for the variance of the estimated instantaneous frequency when the signal is observed in the presence of additive white noise. It does not refer to any formal uncertainty principle associated with our algorithm of §2e. Starting with white noise, we examine the effects of the various steps of the algorithm on the correlation structure of the real and imaginary parts of the resulting complex vectors.

(a) General correlation structure

We can write the DHT matrix defined by the algorithm in Marple (1999) as the circulant

$$
\mathcal{H} = \begin{bmatrix}
0 & q_{N-1}^* & 0 & q_{N-3}^* & 0 & \ldots & 0 & q_1^* \\
q_1^* & 0 & q_{N-1}^* & 0 & q_{N-3}^* & \ldots & q_3^* & 0 \\
0 & q_1^* & 0 & q_{N-1}^* & 0 & \ldots & 0 & q_3^* \\
q_3^* & 0 & q_1^* & 0 & q_{N-1}^* & \ldots & q_5^* & 0 \\
0 & q_3^* & 0 & q_1^* & 0 & \ldots & 0 & q_5^* \\
\vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\
q_{N-1}^* & 0 & \ldots & q_1^* & 0 & \ldots & q_{N-1}^* & 0 \\
q_{N-1}^* & 0 & \ldots & q_1^* & 0 & \ldots & q_{N-1}^* & 0 \\
\end{bmatrix}
$$

(4.1)

where (e.g. Hahn 1996, p. 143) the periodized filter $\{q_l^*\}$ is given by

$$
q_l^* = \frac{2}{N} \sin^2[l\pi/2] \cot[l\pi/N] = \begin{cases} 
\frac{2}{N} \cot[l\pi/N], & l \text{ odd}, \\
0, & l \text{ even}, 
\end{cases}
$$

(4.2)

for $l = 0, \ldots, N-1$. However,

$$
\{q_l^* : l = 0, \ldots, N-1\} \leftrightarrow \{Q_k^* : k = 0, \ldots, N-1\},
$$
where $\leftrightarrow$ denotes ‘Fourier transform pair’ and $Q^*_k$ is given by

$$
Q^*_k = \sum_{l=0}^{N-1} q_l e^{-i2\pi kl/N} = \begin{cases} 
0, & k = 0, \frac{N}{2}, \\
-i, & k = 1, \ldots, \frac{N}{2} - 1, \\
i, & k = \frac{N}{2} + 1, \ldots, N - 1,
\end{cases}
$$

(4.3)

where $k/N$ denotes a Fourier frequency. The definition of $q_l$ means that $\mathcal{H}$ is, importantly, skew-symmetric (i.e. $\mathcal{H} = -\mathcal{H}^T$).

Setting $x = \epsilon$, an $N$-length vector of independent and identically distributed (IID) Gaussian noise (mean zero, variance $\sigma^2$), we firstly create the analytic vector $y = \epsilon + i\mathcal{H}\epsilon = \epsilon + i\tilde{\epsilon}$, say, where $\tilde{\epsilon} \equiv \mathcal{H}\epsilon$. Then,

$$
\text{var}\{\epsilon\} = \sigma^2 I, \quad \text{var}\{\tilde{\epsilon}\} \approx \sigma^2 I, \quad \text{cov}\{\epsilon, \tilde{\epsilon}\} \equiv E\{\epsilon, \tilde{\epsilon}^T\} = \sigma^2 \mathcal{H}^T, \quad \text{cov}\{\epsilon, \epsilon\} = \sigma^2 \mathcal{H}.
$$

The approximation for the second variance follows because of the transfer function structure in equation (4.3)—energy at Fourier frequencies 0 and one-half are removed by the DHT. (Because zero-mean white noise will have no energy at frequency zero, only the latter frequency will matter.) In fact (Olhede & Walden 2004b),

$$
E\{\epsilon_i \tilde{\epsilon}_m\} = \begin{cases} 
\sigma^2 \frac{N-2}{N}, & \text{if } l = m, \\
-\sigma^2 \frac{1 + (-1)^{l-m}}{N}, & \text{otherwise},
\end{cases}
$$

(4.4)

Thus, provided that $N$ is large, then for practical purposes, we can replace the approximation by an equality, and this is done hereafter. Other useful results are

$$
\text{cov}\{\epsilon_i, \tilde{\epsilon}_m\} = \sigma^2 (H^T)_{lm} = -\sigma^2 (H^T)_{ml} = -\text{cov}\{\epsilon_m, \tilde{\epsilon}_l\}
$$

and

$$
\text{cov}\{\epsilon_i, \tilde{\epsilon}_i\} = \sigma^2 (\mathcal{H}^T)_{II} = 0.
$$

Step 2 of the algorithm in §2e consists of the demodulation $d_l = y_l e^{-i2\pi sl}$, $l = 0, \ldots, N-1$. Since $C$ and $S$ are both diagonal matrices they are symmetric (so that e.g. $C = C^T$) and commute with each other. Also, $C$ will commute with any matrix which has all diagonal entries of zero, and similarly for $S$; hence, both $C$ and $S$ will commute with $\mathcal{H}$. Then, with $y = \epsilon + i\tilde{\epsilon}$, we obtain

$$
d = (C - iS)(\epsilon + i\tilde{\epsilon}) = (C\epsilon + S\tilde{\epsilon}) + i(C\tilde{\epsilon} - S\epsilon).
$$

If we now write $d = \Re(d) + i\Im(d)$, we get

$$
\text{var}\{\Re(d)\} = E\{(C\epsilon + S\tilde{\epsilon})(C\epsilon + S\tilde{\epsilon})^T\}
$$

$$
= E\{C\epsilon \epsilon^T C^T + C\epsilon \epsilon^T S^T + S\epsilon \epsilon^T C^T + S\epsilon \epsilon^T S^T\}
$$

$$
= \sigma^2 [CC^T + C\mathcal{H}^T S^T + S\mathcal{H} C^T + S S^T]
$$

$$
= \sigma^2 [CC^T + SS^T] + \sigma^2 [C(\mathcal{H}^T + \mathcal{H}) S^T],
$$

where we have made use of the symmetry and commutative properties of $C$ and $S$. Now, $CC^T + SS^T = I$ and $H^T + H = 0$ since $H$ is skew-symmetric, so we see that $\text{var}\{R(d)\} = \sigma_e^2 I$ and $\text{cov}\{R(d), \Im(d)\} = \mathbb{E}\{(C\epsilon + S\bar{\epsilon})(C\bar{\epsilon} - S\epsilon)^T\}$

$$= \mathbb{E}\{C\epsilon\epsilon^T C^T - C\epsilon\epsilon^T S^T + S\bar{\epsilon}\bar{\epsilon}^T C^T - S\bar{\epsilon}\bar{\epsilon}^T S^T\}$$

$$= \sigma_e^2[C\bar{H}^T C^T - C S^T + S C^T - S \bar{H} S^T] = \sigma_e^2[C\bar{H}^T C^T + \bar{H}^T S^T]$$

$$= \sigma_e^2 \bar{H}^T[CC^T + SS^T] = \sigma_e^2 \bar{H}^T.$$

Therefore, we have shown the covariance structure is unaffected by the demodulation, as would be expected.

Step 3 of the algorithm in §2e creates a new analytic vector $z = d + iHd = d + id$, say, where $d \equiv H d$. We might expect the real and imaginary parts of $z$ to have the same covariance properties as the real and imaginary parts of $y$, because the demodulation has not changed these second-order properties. However, there is an important difference; $z$ is the analytic version of a complex-valued vector, whereas $y$ is the analytic version of a real-valued vector. $z = d + iHd = R(d) - H\Im(d) + i[\Im(d) + H R(d)]$, thus,

$$\text{cov}\{R(z), \Im(z)\} = \mathbb{E}\{(R(d) - H\Im(d))(\Im(d) + H R(d))^T\}$$

$$= \mathbb{E}\{R(d)\Im(d)^T + R(d)R(d)^T H^T - H\Im(d)\Im(d)^T - H\Im(d)R(d)^T H^T\}$$

$$= \sigma_e^2[H^T + H^T - H - H(H H^T)^T] = \sigma_e^2[H^T + \bar{H} (H H^T)] = \sigma_e^2[H^T + \bar{H} (H H^T)].$$

The final line follows from the skew-symmetry of $H$. Now, as shown in appendix B, $H H^T$ is given by the symmetric Toeplitz matrix with the first row

$$\begin{bmatrix} 1 - \frac{2}{N}, 0, -\frac{2}{N}, 0, -\frac{2}{N}, \ldots, -\frac{2}{N}, 0 \end{bmatrix}, \quad (4.5)$$

and as $N \to \infty$, we see that $H H^T \to I$. Hence, for large $N$,

$$\text{cov}\{R(z), \Im(z)\} = 4\sigma_e^2 H^T, \quad (4.6)$$

and similarly, for large $N$,

$$\text{var}\{R(z)\} = 4\sigma_e^2 I \quad \text{and} \quad \text{var}\{\Im(z)\} = 4\sigma_e^2 I. \quad (4.7)$$

Importantly, we see that, apart from the factor of 4, the variances and covariances of the real and imaginary parts of $z$ are the same as for the real and imaginary parts of $y$.

In step 4 of the algorithm, we compute MODWPT projections $\tilde{D}_{j,n} = P_{j,n}z$. Since $P_{j,n}$ is real-valued,

$$\tilde{D}_{j,n} = P_{j,n}z = P_{j,n}R(z) + iP_{j,n}\Im(z) = R(\tilde{D}_{j,n}) + i\Im(\tilde{D}_{j,n}),$$

say. Then,
\[
\text{var}\{\Re(\tilde{D}_{j,n})\} = 4\sigma^2_x P_{j,n} P_{j,n}^T \quad \text{and} \quad \text{var}\{\Im(\tilde{D}_{j,n})\} = 4\sigma^2_x P_{j,n} P_{j,n}^T,
\]
and
\[
\text{cov}\{\Re(\tilde{D}_{j,n}), \Im(\tilde{D}_{j,n})\} = 4\sigma^2_x P_{j,n} \mathcal{H}^T P_{j,n}.
\]

It was shown in Olhede & Walden (2004b) that \( x^T \mathcal{H}^T x = 0 \) for any real-valued vector, \( x \), so that \( (\text{cov}\{\Re(\tilde{D}_{j,n}), \Im(\tilde{D}_{j,n})\})_{ll} = 0 \), \( l=1, \ldots, N \), that is, all diagonal entries of the covariance matrix are zero.

A simple approximate expression for the elements of \( \text{var}\{\Re(\tilde{D}_{j,n})\} \) is derived in equation (C.3), with the same expression for \( \text{var}\{\Im(\tilde{D}_{j,n})\} \). Likewise, in equation (C.2), a simple approximate expression for the elements of \( \text{cov}\{\Re(\tilde{D}_{j,n}), \Im(\tilde{D}_{j,n})\} \), is derived.

Step 5 of the algorithm consists of the reverse demodulation: \( Y_{j,n} = (C + i\delta)\tilde{D}_{j,n} \), which, as we showed for the forward step, will not alter the covariance structure. Hence, if we write \( Y_{j,n} = \Re\{Y_{j,n}\} + i\Im\{Y_{j,n}\} \), then the variances and covariances of \( \Re\{Y_{j,n}\} \) and \( \Im\{Y_{j,n}\} \) are the same as for \( \Re(\tilde{D}_{j,n}) \) and \( \Im(\tilde{D}_{j,n}) \) above.

Hence, starting with \( x = \epsilon \), a zero-mean IID Gaussian noise vector, we have found the (approximate) covariance structure of the complex-valued random vector \( Y_{j,n} \) resulting from the five steps of the algorithm.

\( (b) \) Instantaneous frequency

Since all steps of the algorithm up to and including reverse demodulation consist of linear matrix operations, we know that if our original input signal \( x \) consists of both deterministic and zero-mean white noise components, then the result of the algorithm at step 5 is given by \( Y_{j,n} = Y_{j,n}^{(D)} + Y_{j,n}^{(e)} \), where, henceforth, the superscripts \( D \) and \( e \) indicate the deterministic and noise components, respectively, and \( Y_{j,n}^{(e)} \) has a mean of zero. We can express a component of the deterministic part \( Y_{j,n}^{(D)} \) in terms of its amplitude and phase:
\[
Y_{j,n,l}^{(D)} = A_{j,n,l}^{(D)} e^{i\phi_{j,n,l}^{(D)}}.
\]

However, from equation (2.2), the \( l \) th component of the estimated instantaneous phase vector at level \( j \) frequency band index \( n \) is
\[
\hat{\phi}_{j,n,l} = \tan^{-1}\left[\frac{\Im\{Y_{j,n,l}\}}{\Re\{Y_{j,n,l}\}}\right] = \tan^{-1}\left[\frac{\Im\{Y_{j,n,l}^{(D)}\} + \Im\{Y_{j,n,l}^{(e)}\}}{\Re\{Y_{j,n,l}^{(D)}\} + \Re\{Y_{j,n,l}^{(e)}\}}\right],
\]
and a first-order Taylor expansion about \( (\Re\{Y_{j,n,l}^{(D)}\}, \Im\{Y_{j,n,l}^{(D)}\}) \) gives
\[
\hat{\phi}_{j,n,l} \approx \tan^{-1}\left[\frac{\Im\{Y_{j,n,l}^{(D)}\}}{\Re\{Y_{j,n,l}^{(D)}\}}\right] + \frac{\left[\Im\{Y_{j,n,l}^{(D)}\}\Re\{Y_{j,n,l}^{(e)}\} - \Im\{Y_{j,n,l}^{(e)}\}\Re\{Y_{j,n,l}^{(D)}\}\right]}{\Re^2\{Y_{j,n,l}^{(D)}\} + \Im^2\{Y_{j,n,l}^{(D)}\}}. \tag{4.9}
\]
From equation (4.8), we know that $[A_{j,n,l}^{(D)}]^2 = \Re\{Y_{j,n,l}^{(D)}\} + \Im\{Y_{j,n,l}^{(D)}\}$. Since the variances and covariances of $\Re\{Y_{j,n}^{(e)}\}$ and $\Im\{Y_{j,n}^{(e)}\}$ are the same as for $\Re\{\tilde{\phi}_{j,n}\}$ and $\Im\{\tilde{\phi}_{j,n}\}$, we use equations (C.2) and (C.3) to see that the variance of the estimated phase—since there is now additive noise—is

$$\var\{\hat{\phi}_{j,n,l}\} \approx \frac{4\sigma_s^2}{2[A_{j,n,l}^{(D)}]^2}.$$  

The estimated instantaneous frequency $\hat{\nu}_{j,n,l}$ is got by substituting the estimated phase into equation (2.3). Hence, its variance takes the form

$$\var\{\hat{\nu}_{j,n,l}\} = \frac{1}{4\pi^2} \text{cov}\left\{ \sum_{m=-2}^{2} r_m \hat{\phi}_{j,n,l-m}, \sum_{k=-2}^{2} r_k \hat{\phi}_{j,n,l-k}\right\}$$  

$$= \frac{1}{4\pi^2} \sum_{m=-2}^{2} \sum_{k=-2}^{2} r_m r_k \text{cov}\{\hat{\phi}_{j,n,l-m}, \hat{\phi}_{j,n,l-k}\}.$$  

(4.10)

Now, using equation (4.9), $\text{cov}\{\hat{\phi}_{j,n,l-m}, \hat{\phi}_{j,n,l-k}\}$ can be expressed as

$$E\{[\Re\{Y_{j,n,l-m}^{(D)}\}\Im\{Y_{j,n,l-m}^{(D)}\} - \Im\{Y_{j,n,l-m}^{(D)}\}\Re\{Y_{j,n,l-m}^{(D)}\}] [\Re\{Y_{j,n,l-k}^{(D)}\}\Im\{Y_{j,n,l-k}^{(D)}\} - \Im\{Y_{j,n,l-k}^{(D)}\}\Re\{Y_{j,n,l-k}^{(D)}\}]\}.$$  

Thus, $\text{cov}\{\hat{\phi}_{j,n,l-m}, \hat{\phi}_{j,n,l-k}\}$ is approximately

$$\{\Re\{Y_{j,n,l-m}^{(D)}\}\Re\{Y_{j,n,l-k}^{(D)}\}\Xi_{j,n,m-k} - \Im\{Y_{j,n,l-m}^{(D)}\}\Im\{Y_{j,n,l-k}^{(D)}\} \Re\{Y_{j,n,l-m}^{(D)}\} \Im\{Y_{j,n,l-k}^{(D)}\}\}$$  

$$- \Im\{Y_{j,n,l-m}^{(D)}\}\Re\{Y_{j,n,l-k}^{(D)}\} Y_{j,n,m-k}$$  

$$+ \Re\{Y_{j,n,l-m}^{(D)}\}\Im\{Y_{j,n,l-k}^{(D)}\} \Xi_{j,n,m-k} = \{[A_{j,n,l-m}^{(D)}]^2[A_{j,n,l-k}^{(D)}]^2\},$$

where the symmetric autocovariance sequence, $\{\Xi_{j,n,\tau}\}$, and approximate skew-symmetric cross-covariance sequence, $\{Y_{j,n,\tau}\}$, are defined in equations (C.2) and (C.3). However,

$$\Re\{Y_{j,n,l-m}^{(D)}\}\Re\{Y_{j,n,l-k}^{(D)}\} = A_{j,n,l-m}^{(D)} A_{j,n,l-k}^{(D)} \cos(\phi_{j,n,l-m}^{(D)}) \cos(\phi_{j,n,l-k}^{(D)}),$$

and similarly for the other terms. Our approximate form for $\text{cov}\{\hat{\phi}_{j,n,l-m}, \hat{\phi}_{j,n,l-k}\}$ becomes

$$\{[\cos(\phi_{j,n,l-m}^{(D)}) \cos(\phi_{j,n,l-k}^{(D)}) + \sin(\phi_{j,n,l-m}^{(D)}) \sin(\phi_{j,n,l-k}^{(D)})] \Xi_{j,n,m-k}$$  

$$+ \{\sin(\phi_{j,n,l-m}^{(D)}) \cos(\phi_{j,n,l-k}^{(D)}) - \cos(\phi_{j,n,l-m}^{(D)}) \sin(\phi_{j,n,l-k}^{(D)})\} Y_{j,n,m-k}\} /$$  

$$\{[A_{j,n,l-m}^{(D)}]^2[A_{j,n,l-k}^{(D)}]^2\}$$

$$= \{[\cos(\phi_{j,n,l-m}^{(D)} - \phi_{j,n,l-k}^{(D)})] \Xi_{j,n,m-k} + \{\sin(\phi_{j,n,l-m}^{(D)} - \phi_{j,n,l-k}^{(D)})\} Y_{j,n,m-k}\} /$$  

$$\{[A_{j,n,l-m}^{(D)}]^2[A_{j,n,l-k}^{(D)}]^2\}.$$  

---

S. Olhede and A. T. Walden

Finally, our approximation for \( \text{var}\{\hat{\nu}_{j,n,l}\} \) is thus given by
\[
\sum_{m,k=-2}^{2} r_m r_k \frac{[\{\cos(\phi_{j,n,l-m}^{(D)} - \phi_{j,n,l-k}^{(D)})\} \Sigma_{j,n,m-k} + \{\sin(\phi_{j,n,l-m}^{(D)} - \phi_{j,n,l-k}^{(D)})\} \Gamma_{j,n,m-k}]}{4\pi^2 A_{j,n,l-m}^{(D)} A_{j,n,l-k}^{(D)}}.
\]

When calculating this expression with real data, the sequences of amplitudes, \( \{A_{j,n,l}^{(D)}\} \), and phases, \( \{\phi_{j,n,l}^{(D)}\} \), are unknown, but these can be replaced by their estimates.

In order to show the validity of the above variance expression, consider the linear chirp \( x_l = \sin(a_1 t^2 + a_2 t), \ l = 0, \ldots, 1023, \) with \( a_1 = 0.02\pi/1024 \) and \( a_2 = 0.36\pi \). This has an instantaneous frequency increasing from 0.18 to 0.2 over the 1024 time-points. White Gaussian noise was added to achieve a signal-to-noise variance ratio of 10. A level \( J_0=2 \) MODWPT was used; the signal lies in band \( n = 1 \), frequency band \([0.125,0.25] \). \( \text{var}\{\hat{\nu}_{2,1,l}\} \) was calculated using equation (4.11) and the known amplitudes \( \{A_{2,1,l}^{(D)}\} \) and phases \( \{\phi_{2,1,l}^{(D)}\} \) and is shown in figure 3a. Figure 3b shows the

Figure 3. (a) Theoretical standard deviation from equation (4.11) of estimated instantaneous frequency; (b) the mean standard deviation from equation (4.11) from 200 independent repeat simulations using quantities estimated from the noisy data; (c) comparison of (b) (dashed) with the empirical standard deviation computed from 200 independent repeat simulations of the estimation of instantaneous frequency. Calculations used the MODWPT with \( J_0=2 \) and 18-point Fejér–Korovkin wavelet filters.
mean standard deviation obtained at each time from 200 independent repeat simulations, again using equation (4.11), but with the amplitudes \{A_{2,1,l}^{(D)}\} and phases \{\phi_{2,1,l}^{(D)}\} estimated from the noisy data of each simulation. Finally, figure 3c compares the result shown in figure 3b with the empirical standard deviation computed from the 200 independent repeat simulations of the estimation of instantaneous frequency at each time. It can be seen that there is good agreement between all three cases, supporting the validity of equation (4.11).

When calculating equation (4.11) with real data, not only must the sequences of amplitudes, \{A_{j,n,l}^{(D)}\}, and phases, \{\phi_{j,n,l}^{(D)}\}, be estimated, but so too must the variance of the white noise, \(\sigma_e^2\), occurring in the formula for \(\{\Xi_{j,n,\tau}\}\) in equation (C 3) and in the formula for \(\{\Upsilon_{j,n,\tau}\}\) in equation (C 2). This variance can be estimated from the MODWPT details in an appropriate signal-free level and band. For example, for figure 3b, we used \(j = J_0 = 2\) and \(n = 3\), which corresponds to the highest frequency band. When incorporating the demodulation and reverse demodulation steps we must proceed with caution: energy moved out of the positive frequency region of time-frequency into the negative region (via wraparound) by the demodulation in step 2 of the algorithm in §2c will be eliminated in the creation of the new analytic vector in step 3. Of course, we assume that the demodulation is chosen so that the signal is not eliminated in this way, but some noise will inevitably be, so that when estimating \(\sigma_e^2\) in this case, care should be taken not to choose MODWPT details in a signal-free region where the noise has been affected by the demodulation.

5. Summary

We have described a flexible approach for the time-frequency analysis of multicomponent signals. It involves the use of analytic vectors and demodulation. The demodulated analytic signal is projected onto the time-frequency plane so that, as closely as possible, each component contributes exclusively to a different ‘tile’ in a wavelet packet tiling of the time-frequency plane, and at any time the contribution to each tile definitely comes from no more than one component. A single reverse demodulation is then applied to all projected components. Two examples of how to choose the form of the demodulation sequence \(s_l\) were presented in some detail within the context of the real and simulated data examples. Following the above procedures, the instantaneous frequency of each component in each tile is well defined and may be calculated, followed by the Hilbert energy spectrum.

In order to better understand the effect of additive noise, the approximate variance of the estimated instantaneous frequency in any tile has been formulated by starting with pure noise and studying its evolving covariance structure through each step of the algorithm. The validity and practical utility of the resulting expression was demonstrated by comparing the theoretical standard deviation with all parameters taken as known, with its mean over 200 simulations with unknown quantities estimated from the noisy data, and with the empirical standard deviation computed from 200 simulations of the estimation of instantaneous frequency.

The site http://www.ma.imperial.ac.uk/statistics/research/wavelets/ contains MATLAB code for our algorithm.
We wish to show that the matrices \( P_{j,n} = \tilde{W}_{j,n}^T \tilde{W}_{j,n} \) and \( \mathcal{H} \) commute; from equation (2.1), we can write \( \tilde{W}_{j,n} \) as

\[
\tilde{W}_{j,n} = \begin{bmatrix}
\tilde{u}_{j,n,0}^* & \tilde{u}_{j,n,1}^* & \tilde{u}_{j,n,2}^* & \tilde{u}_{j,n,3}^* & \cdots & \tilde{u}_{j,n,N-1}^* \\
\tilde{u}_{j,n,1}^* & \tilde{u}_{j,n,0}^* & \tilde{u}_{j,n,2}^* & \tilde{u}_{j,n,3}^* & \cdots & \tilde{u}_{j,n,N-1}^* \\
\tilde{u}_{j,n,2}^* & \tilde{u}_{j,n,1}^* & \tilde{u}_{j,n,0}^* & \tilde{u}_{j,n,3}^* & \cdots & \tilde{u}_{j,n,N-1}^* \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\tilde{u}_{j,n,N-1}^* & \tilde{u}_{j,n,N-2}^* & \tilde{u}_{j,n,N-3}^* & \tilde{u}_{j,n,N-4}^* & \cdots & \tilde{u}_{j,n,0}^* \\
\end{bmatrix}
\]  

(see Percival & Walden 2000, p. 171).

To show that the matrices \( P_{j,n} = \tilde{W}_{j,n}^T \tilde{W}_{j,n} \) and \( \mathcal{H} \) commute, we firstly define the circular shift (orthonormal) transform matrix \( \mathcal{T} \). Then, with \( \mathbf{x} = [x_0, \ldots, x_{N-1}]^T \), we have \( \mathcal{T} \mathbf{x} = [x_{N-1}, x_0, \ldots, x_{N-2}]^T \). Now, \( \mathcal{T}^2 = \mathcal{T} \mathcal{T} \) and so on. We can thus write \( \mathcal{H} \) in equation (4.1) as \( \mathcal{H} = \sum_{m=0}^{N-1} q_m^* \mathcal{T}^m \), where \( q_m^* = 0 \) for \( m \) even. Similarly, \( \tilde{W}_{j,n} \) in (A 1) can be written as \( \tilde{W}_{j,n} = \sum_{m=0}^{N-1} \tilde{u}_{j,n,k} \mathcal{T}^m \) and \( \tilde{W}_{j,n}^T = \sum_{m=0}^{N-1} \tilde{u}_{j,n,k}^* (\mathcal{T}^T)^k = \sum_{m=0}^{N-1} \tilde{u}_{j,n,k}^* \mathcal{T}^{-k} \), given that \( \mathcal{T}^T = \mathcal{T}^{-1} \) because \( \mathcal{T} \) is orthonormal. Hence,

\[
\mathcal{H} \tilde{W}_{j,n}^T \tilde{W}_{j,n} = \sum_{m=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} q_m^* \tilde{u}_{j,n,k}^* \tilde{u}_{j,n,l} \mathcal{T}^{-k-l+m} = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} \tilde{u}_{j,n,k}^* \tilde{u}_{j,n,l} q_m^* \mathcal{T}^{-k-l+m}
\]

as required.

### Appendix B

To show that \( \mathcal{H} \mathcal{H}^T \) has the form stated in equation (4.5), it is sufficient to consider the dot product of the first row of \( \mathcal{H} \) with circularly shifted versions of other rows. Now, \( q^* \star q_l^* = \sum_{n=0}^{N-1} q_n^* q_{(n+l) \mod N} \) and \( \{ q^* \star q_l^* \} \leftrightarrow \{ |Q_k^*|^2 \} \equiv \mathcal{Q}_k \).

The inverse Fourier transform relationship yields \( q^* \star q_l^* = \frac{1}{N} \sum_{k=0}^{N-1} \mathcal{Q}_k^* e^{2\pi i kl/N} \). However, from equation (4.3), we know that

\[
\mathcal{Q}_k^* = \begin{cases} 
0, & k = 0, \frac{N}{2}, \\
1, & k = 1, \ldots, \frac{N}{2} - 1, \text{ and } k = \frac{N}{2} + 1, \ldots, N - 1.
\end{cases}
\]
Also,

$$\frac{1}{N} \sum_{k=0}^{N-1} e^{i2\pi lk/N} = \begin{cases} 1, & l = 0, \\ 0, & l = 1, \ldots, N-1, \end{cases}$$

so that

$$\sum_{n=0}^{N-1} q_n^* q_{(n+l)\text{mod } N} + \frac{1}{N} (1 + [-1]^l) = \begin{cases} 1, & l = 0, \\ 0, & l = 1, \ldots, N-1, \end{cases}$$

namely,

$$\sum_{n=0}^{N-1} q_n^* q_{(n+l)\text{mod } N} = \begin{cases} 1 - \frac{2}{N}, & l = 0, \\ 0, & l = 1, 3, 5, \ldots, N-1, \\ -\frac{2}{N}, & l = 2, 4, 6, \ldots, N-2. \end{cases}$$

The form of $\mathcal{H}\mathcal{H}^T$ in equation (4.5) then follows immediately.

### Appendix C

Here, we derive simple approximate expressions for the elements of $\text{var}\{\Re(\tilde{D}_{j,n})\}$, $\text{var}\{\Im(\tilde{D}_{j,n})\}$ and $\text{cov}\{\Re(\tilde{D}_{j,n}), \Im(\tilde{D}_{j,n})\}$. Since $P_{j,n}$ is real-valued we know that

$$\tilde{D}_{j,n} = P_{j,n} z = P_{j,n} \Re(z) + i P_{j,n} \Im(z) = \Re(\tilde{D}_{j,n}) + i\Im(\tilde{D}_{j,n}).$$

We shall make use of the shorthand $a_{j,n} = \Re(\tilde{D}_{j,n})$; $b_{j,n} = \Im(\tilde{D}_{j,n})$. The variances and covariances of $\Re(z)$ and $\Im(z)$ are specified in equations (4.6) and (4.7).

Now, consider $\text{cov}\{a_{j,n,l}, b_{j,n,m}\}$.

Let $r^T$ be the row vector of $P_{j,n}$ that yields $a_{j,n,l}$ in $a_{j,n} = P_{j,n} \Re(z)$ and let $v^T$ be the row vector of $P_{j,n}$ that yields $b_{j,n,m}$ in $b_{j,n} = P_{j,n} \Im(z)$. Now, $\text{cov}\{\Re(z), \Im(z)\} = 4\sigma_z^2 \mathcal{H}^T$ (see equation (4.6)) and because $\mathcal{H}^T$ is a circulant matrix, $\mathcal{H}^T = \mathcal{F}^H A \mathcal{F}$ (e.g. Percival & Walden 2000, p. 441).

The $(l,m)$th element of the orthonormal DFT (ODFT) matrix $\mathcal{F}$ is given by \(\exp(-i2\pi lm/N)/\sqrt{N}\), for $0 \leq l, m \leq N-1$, $^H$ denotes Hermitian transpose, and $A$ is a diagonal matrix with diagonal elements $\{Q_k^e \colon k = 0, \ldots, N-1\}$, the DFT of the first row of $\mathcal{H}^T$. Then,

$$\text{cov}\{a_{j,n,l}, b_{j,n,m}\} = 4\sigma_z^2 r^T \mathcal{H}^T v = 4\sigma_z^2 r^T \mathcal{F}^H A \mathcal{F} v = 4\sigma_z^2 (\mathcal{F} r)^H A \mathcal{F} v.$$

Now, $\mathcal{F} r$ is the ODFT of $r$ which is related to the DFT, say, $\{R_k\}$ of $r$ via $\mathcal{F} r = (1/\sqrt{N})[R_0, \ldots, R_{N-1}]^T$, and similarly for $\mathcal{F} v$. Hence,

$$\text{cov}\{a_{j,n,l}, b_{j,n,m}\} = \frac{4\sigma_z^2}{N} [R_0^*, R_1^*, \ldots, R_{N-1}^*] \begin{bmatrix} Q_0^* & 0 & \cdots & 0 \\ 0 & Q_1^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q_{N-1}^* \end{bmatrix} \begin{bmatrix} V_0 \\ V_1 \\ \vdots \\ V_{N-1} \end{bmatrix}$$

$$= \frac{4\sigma_z^2}{N} \sum_{k=0}^{N-1} R_k^* Q_k V_k.$$
With \( U_{j,n,k}^* = \sum_{l=0}^{N-1} \tilde{u}_{j,n,l}^* e^{-i2\pi kl/N} \), we have \( R_k^* = |U_{j,n,k}^*|^2 e^{i2\pi kl/N} \) and \( V_k = |U_{j,n,k}^*|^2 e^{-i2\pi km/N} \), so that

\[
\text{cov}\{a_{j,n,l}, b_{j,n,m}\} = \frac{4\sigma^2}{N} \sum_{k=0}^{N-1} |U_{j,n,k}^*|^4 Q_k e^{i2\pi k(l-m)/N}. \tag{C 1}
\]

However, \( U_{j,n,k}^* = U_{j,n}(k/N) \), where \( U_{j,n}(f) = \sum_{l=0}^{L_j-1} \tilde{u}_{j,n,l} e^{-i2\pi lf} \), we recall that \( \{\tilde{u}_{j,n,l}\} \) is a \( J \)th level and \( N \)th band MODWPT wavelet packet filter of length \( L_j = (2^J - 1)(L - 1) + 1 \) while \( \tilde{u}_{j,n,l} \) is its periodized version. Also, \( Q_k = Q(k/N) \) (Percival & Walden 2000, p. 506), where \( Q(f) = \sum_{l=-\infty}^{\infty} q_l e^{-i2\pi lf} \) is the periodic function

\[
Q(f) = -i \text{sgn}_p(f) = \begin{cases} 0, & \text{if } f = 0, 1/2, \\ -i, & \text{if } 0 < f < 1/2, \\ i, & \text{if } -1/2 < f < 0, 
\end{cases}
\]

and \( \{q_l\} \) is the non-periodized Hilbert transform filter given by

\[
q_l = \int_{-1/2}^{1/2} -i \text{sgn}_p(f) e^{i2\pi f l} df = \begin{cases} \frac{2 \sin^2(l\pi/2)}{l\pi}, & \text{if } l \text{ odd}, \\ 0, & \text{if } l \text{ even}, 
\end{cases}
\]

so that \( q_l^* = \sum_{n=-\infty}^{\infty} q_{l+nN} \). It then follows from equation (C 1) that

\[
\text{cov}\{a_{j,n,l}, b_{j,n,m}\} \approx 4\sigma^2 \int_{-1/2}^{1/2} |U_{j,n}(f)|^4 Q(f) e^{i2\pi f(l-m)} df. \]

However,

\[
|U_{j,n}(f)|^2 \approx \begin{cases} 1, & \text{if } f \in \left\{-\left(\frac{n+1}{2^j+1}, \frac{n}{2^j+1}\right) \cup \left(\frac{n}{2^j+1}, \frac{n+1}{2^j+1}\right)\right\}, \\ 0, & \text{otherwise}, 
\end{cases}
\]

so that

\[
\text{cov}\{a_{j,n,l}, b_{j,n,m}\} \approx \frac{4\sigma^2}{2\pi(l-m)} \left| e^{i2\pi f(l-m)} \right|^{n/2^j+1} - \left| e^{i2\pi f(l-m)} \right|^{(n+1)/2^j+1} \]

\[
= \begin{cases} 0, & l = m, \\ 4\sigma^2 \frac{\left| \cos(n\pi(l-m)/2^j) - \cos((n+1)\pi(l-m)/2^j) \right|}{\pi(l-m)} & l \neq m, 
\end{cases}
\]

\( \equiv r_{j,n,l-m} \). \tag{C 2}

This is skew-symmetric in \( l \) and \( m \).

We can determine \( \text{cov}\{a_{j,n,l}, a_{j,n,m}\} \) (or \( \text{cov}\{b_{j,n,l}, b_{j,n,m}\} \)) by noting from equation (4.7) that the appropriate circulant matrix in this case is simply \( \mathbf{I} \); the
DFT of its first row is unity, so that

\[
\text{cov}\{a_{j,n,l}, a_{j,n,m}\} \approx 4\sigma^2 \int_{-1/2}^{1/2} |U_{j,n}(f)|^4 e^{2\pi i (l-m) df}
\]

\[
= \begin{cases} 
4\sigma^2/2^i, & l = m, \\
4\sigma^2 \frac{\sin[(n+1)\pi(l-m)/2^i] - \sin[n\pi(l-m)/2^i]}{\pi(l-m)}, & l \neq m,
\end{cases}
\]

\[
\equiv \Xi_{j,n,l-m}.
\]  

(C 3)

with an identical result for \(\text{cov}\{b_{j,n,l}, b_{j,n,m}\}\); note this is symmetric in \(l\) and \(m\).

References


Analysis of multicomponent signals


