Polarization phase relationships via multiple Morse wavelets — Part I: Fundamentals

BY S. OLHEDE AND A.T. WALDEN

Department of Mathematics, Imperial College of Science, Technology and Medicine, 180 Queen's Gate, London SW7 2BZ, UK.

The purpose of this paper is to provide a detailed analysis of the use of multiple generalized Morse wavelets for polarization analysis of multi-component recordings where phase relationships between components of the signal can be transient. Special attention will be given to the case of three components. The use of complex and analytic wavelets enables detection of coherent motion with elliptical polarization because of sensitivity to phase shifts among components. We adopt a singular value decomposition approach. The principal polarization is given by the first eigenvector of the multiple wavelet multivariate scalogram at scale a and time b . Having described in detail a deterministic signal plus random noise model for the recorded data, we show that for (a, b) in a domain Ω , we can approximate the phase between components of the original multi-component signal via the estimated phase of the components of the first right-singular vector; detailed results on the accuracy of the approximation are given. The exact form of the domain Ω , which depends on the transient nature of the phase relationship, is described.

Keywords: multi-component signals; multiple wavelets; non-stationary time series; phase; polarization; singular value decomposition; statistical analysis

1. Introduction

Polarization analysis aids in identifying and analysing waves present in multi-component signals. Recently interest has focussed on using complex-valued wavelet methods in polarization studies. In a solid-earth geophysical context Lilly and Park (1995) introduced the idea of studying elliptical polarization (where there is a phase lag between components) by forming multiple complex-valued wavelets from contiguous odd and even real-valued Slepian wavelets, leading to multiple wavelet time-varying polarization analysis for seismic records. Various aspects of their methodology have been used in practice (e.g., Bear and Pavlis 1997; Zandrea *et al.* 2000). In a solar physics context, Lucek and Balogh (1998) used a single Morlet wavelet (a well-known complex-valued wavelet) to look for polarized waves in the interplanetary magnetic field observed by the Ulysses satellite upstream of a shock in the interplanetary medium.

The use of multiple orthogonal wavelets, instead of a single wavelet, enables use of an averaging scheme over the orthogonal wavelets to reduce variance. Such an approach has proven very advantageous in the field of spectrum analysis of stationary stochastic or noisy signals where Thomson (1982) proposed the use of multiple

orthogonal data tapers, each giving rise to a different spectrum estimate; these spectrum estimates are then averaged, and the resulting estimate is more interpretable due to the reduced variance. For estimation of the scalogram (modulus-squared of continuous wavelet transform) of stochastic or noisy non-stationary signals, the same approach can be used — several orthogonal wavelets may be used to create a set of different scalograms, which are averaged together to produce a low variance scalogram estimate. Such scalograms will be the key ingredient in wavelet-based polarization analysis.

The orthogonality of the above multiple functions follows because they arise as the eigenfunctions of a localisation operator. All localisation operators are associated with a domain $\mathcal{D}$ in the time-frequency plane $((t, f) \in \mathbb{R}^2)$ over which we are trying to maximally concentrate the signal.

A classical set of multiple orthogonal functions, the prolate spheroidal wave functions, are the eigenfunctions associated with the operator which truncates a signal in time, and then in frequency, the domain in time-frequency space being a rectangle centred at time-zero frequency-zero. In discrete time the resulting orthogonal sequences are often known as Slepian sequences and arise from maximizing the energy of a finite-length discrete-time signal in a frequency band $|f| \leq f_w$. Lilly and Park (1995) looked at the problem of maximizing the energy of a finite-length discrete-time signal in a frequency band $|f - f_c| \leq f_w$ about $\pm f_c$. This is similar to the constraint leading to Slepian sequences, except importantly, the frequency band is centred on $\pm f_c$ rather than zero. Let us denote the resulting k th eigenvector (ordered by corresponding eigenvalue) by $\psi_k(f_c)$. Lilly and Park (1995) ingeniously constructed artificial complex-valued Slepian pseudo-wavelets from these real eigenvectors using

$$\tilde{\psi}_k^{(S)}(f_c) = \frac{1}{\sqrt{2}}[\psi_{2k}(f_c) + i\psi_{2k+1}(f_c)], \quad (1.1)$$

as the order- k complex Slepian pseudo-wavelet for $k = 0, 1, 2$ etc. up to some well-defined limit. These multiple complex pseudo-wavelets may be used for wavelet analysis at frequency f_c . Note they are defined discretely and may be applied using a discrete convolution as set out in Lilly and Park (1995).

To treat time and frequency *simultaneously* Daubechies (1988) and Daubechies and Paul (1988) introduced new localisation operators which can be used in the time-frequency plane. They provided solutions for two special cases: when the domain consists of an ellipse the resulting orthogonal eigenfunctions are scaled versions of the Hermite functions, and when the domain consists of a ‘Morse region’ (defined later in equation (2.1)) the orthogonal eigenfunctions obtained are known as the complex-valued generalized Morse wavelets (further developed in detail in Olhede and Walden (2002a)).

Instead of using the multiple complex Slepian pseudo-wavelets for polarization analysis, we use here the multiple generalized complex Morse wavelets. In §2 we give the important properties of these wavelets. In §3 we discuss the construction of the multiple wavelet multivariate scalogram $\hat{\mathbf{S}}$ at scale a and time b . Its eigenvalues and eigenvectors are examined in §4; the eigenvector corresponding to the largest eigenvalue provides the polarization vector. A model for the observed multi-component data in terms of a deterministic signal of general form plus isotropic noise is given detailed consideration in §5. The accuracy of estimation of polarization phase from polarization vector components is treated in §6. Importantly, it is shown that for

(a, b) in a fully described domain Ω , we can approximate the transient phase between components of the original multi-component signal via the estimated polarization phase. The superiority of the Morse wavelets over the Slepian pseudo-wavelets for phase determination is also demonstrated. Detailed technical results are set out in the Appendices, including a careful treatment of approximation orders.

2. The generalized Morse wavelets

In this section we briefly summarize relevant results given in Olhede and Walden (2002a), and also provide some additional information where appropriate.

(a) Domain for Morse concentration operator

For the Morse wavelets we start with a concentration operator $\mathcal{P}_{\mathcal{D}_{C,\beta,\gamma}}$, dependent on three parameters C, β and γ . The domain $\mathcal{D} \equiv \mathcal{D}_{C,\beta,\gamma}$ of the time-frequency plane over which the operator $\mathcal{P}_{\mathcal{D}_{C,\beta,\gamma}}$ can be characterised is given by

$$\mathcal{D}_{C,\beta,\gamma} = \left\{ (t, f) \in \mathbb{R}^2 : \left(\frac{t}{C_2} \right)^2 \left(\frac{C_1}{|2\pi f|} \right)^{2\gamma-2} + \left(\frac{C_1}{|2\pi f|} \right)^{2\gamma} + 1 \leq 2C \left(\frac{C_1}{|2\pi f|} \right)^\gamma \right\}, \quad (2.1)$$

where $C_1 = 2^{-\frac{1}{\gamma}} \Gamma(r + \frac{1}{\gamma}) / \Gamma(r)$, $C_2 = \beta \gamma^{-1} 2^{\frac{1}{\gamma}} \Gamma(r - \frac{1}{\gamma}) / \Gamma(r)$, $\Gamma(\cdot)$ denotes the Gamma function and $r = (2\beta + 1)/\gamma$. The region $\mathcal{D}_{C,\beta,\gamma}$ consists of two parts symmetrically placed about the $f = 0$ line. The area of the domain, $|\mathcal{D}_{C,\beta,\gamma}|$, is given explicitly in Olhede and Walden (2002a).

The energy concentration corresponding to the k th-order generalized Morse eigenfunction is given by the square of the k th eigenvalue of the concentration operator; it depends on the area, $|\mathcal{D}_{C,\beta,\gamma}|$, of the domain in the time-frequency plane. A key finding in Olhede and Walden (2002a) is that if $\gamma > 1$ high energy concentrations can be obtained, exceeding that of the Hermite eigenfunctions, which are themselves considered to have good time-frequency concentration (e.g., Vakman 1967; Çakrak and Loughlin 2001). An example of this good time-frequency concentration is shown in figure 1. The left panel of figure 1 shows the eigenvalues of the operator for orders $k = 0, \dots, 4$ and areas $|\mathcal{D}_{C,\beta,\gamma}|$, of 100 (dashed) and 250 (solid) with $\beta = 8, \gamma = 3$ and C determined by the stated area; the corresponding domains $\mathcal{D}_{C,\beta,\gamma}$ of the operators for the two cases are shown in the right panel.

For $f > 0$ the minimum, $f_{\triangleright}$, and maximum, $f_{\triangleleft}$, frequency bounds of $\mathcal{D}_{C,\beta,\gamma}$ must occur at $t = 0$ and are given by

$$\left. \begin{matrix} f_{\triangleright} \\ f_{\triangleleft} \end{matrix} \right\} = C_1 / \left\{ 2\pi [C \pm \sqrt{(C^2 - 1)}]^\frac{1}{\gamma} \right\}.$$

To find the minimum, $t_{\triangleright}$, and maximum, $t_{\triangleleft}$, time bounds of $\mathcal{D}_{C,\beta,\gamma}$ we differentiate the domain boundary function specified by (2.1) with respect to frequency and setting to zero gives the positive frequency

$$f^* = \frac{1}{\pi} 2^{\frac{1}{\gamma}-1} C_1 \left(C(2 - \gamma) + \sqrt{4(\gamma - 1) + C^2(\gamma - 2)^2} \right)^{-\frac{1}{\gamma}},$$

Figure 1. Left: Morse operator eigenvalues of orders $k = 0, \dots, 4$ for $\beta = 8, \gamma = 3$ for areas 100 (dashed) and 250 (solid). Right: Matching domains of the operators in the two cases. For the larger domain the dotted horizontal lines show the (positive) frequency extrema f_+ and f_- while the vertical dotted lines show the time extrema t_+ and t_- .

which may be substituted in (2.1) to give

$$\left. \begin{matrix} t_+ \\ t_- \end{matrix} \right\} = \mp C_2 \left(\frac{C_1}{2\pi f^*} \right)^{1-\gamma} \sqrt{2C \left(\frac{C_1}{2\pi f^*} \right)^\gamma - \left(\frac{C_1}{2\pi f^*} \right)^{2\gamma} - 1}.$$

The frequency and time extremes are illustrated in figure 1.

(b) The generalized Morse wavelets

The k th eigenvalue of the concentration operator has multiplicity two with associated Hermitian eigenfunctions $\psi_{k;\beta,\gamma}^+(t)$, an analytic wavelet, and $\psi_{k;\beta,\gamma}^-(t)$, an anti-analytic wavelet, both of norm unity. By analytic we mean that the imaginary part of $\psi_{k;\beta,\gamma}^+(t)$ is the Hilbert transform of the real part; $\psi_{k;\beta,\gamma}^-(t)$ is the complex conjugate of $\psi_{k;\beta,\gamma}^+(t)$. (Details on computation of $\psi_{k;\beta,\gamma}^+(t)$ may be found in Olhede and Walden 2002a). The Fourier transforms of these functions, $\Psi_{k;\beta,\gamma}^+(f)$ and $\Psi_{k;\beta,\gamma}^-(f)$, are real-valued on the positive and negative frequency axes, respectively, and zero elsewhere. They are given by (Daubechies and Paul 1988, p. 679):

$$\Psi_{k;\beta,\gamma}^+(f) = \sqrt{2} A_{k;\beta,\gamma} (2\pi f)^\beta e^{-(2\pi f)^\gamma} L_k^c(2[2\pi f]^\gamma), \quad (2.2)$$

for $f > 0$, and zero otherwise, and

$$\Psi_{k;\beta,\gamma}^-(f) = \sqrt{2} A_{k;\beta,\gamma} (2\pi |f|)^\beta e^{-(2\pi |f|)^\gamma} L_k^c(2[2\pi |f|]^\gamma),$$

for $f < 0$, and zero otherwise. Here $c = r-1$ and $A_{k;\beta,\gamma} = \sqrt{[\pi\gamma 2^r \Gamma(k+1)/\Gamma(k+r)]}$, and $L_k^c(\cdot)$ denotes the generalized Laguerre polynomial:

$$L_k^c(x) = \sum_{m=0}^k (-1)^m \frac{\Gamma(k+c+1)}{\Gamma(c+m+1)\Gamma(k-m+1)} \frac{x^m}{m!}.$$

Figure 2. Left: Even (solid) and odd (dashed) generalized Morse wavelets for $\beta = 8, \gamma = 3$, for $k = 0$ (top) to $k = 4$ (bottom). The dotted lines mark the time extrema $t_{\triangleright}$ and $t_{\triangleleft}$ for the domain with area 250 as seen in figure 1. Right: Corresponding frequency functions $\Psi_{k;\beta,\gamma}^+(f)$. The dotted lines mark the frequency extrema $f_{\triangleright}$ and $f_{\triangleleft}$ for the domain with area 250 as seen in figure 1.

The k th-order complex and analytic generalized Morse wavelet (eigenfunction) can be written in terms of two real Morse wavelets, one even, $\psi_{k;\beta,\gamma}^{(e)}(t)$, and one odd, $\psi_{k;\beta,\gamma}^{(o)}(t)$,

$$\psi_{k;\beta,\gamma}^+(t) = \frac{1}{\sqrt{2}}[\psi_{k;\beta,\gamma}^{(e)}(t) + i\psi_{k;\beta,\gamma}^{(o)}(t)]. \quad (2.3)$$

The even and odd wavelets are shown for $k = 0, \dots, 4$ in the left panel of figure 2 for $\beta = 8, \gamma = 3$; their corresponding Fourier transforms $\Psi_{k;\beta,\gamma}^+(f)$ are shown in the right panel. Figure 1 indicated that for $\beta = 8, \gamma = 3$ the first five generalized Morse eigenfunctions are almost perfectly concentrated within the domain of area 250 (the squared eigenvalues are almost exactly one). Hence the corresponding time extrema $t_{\triangleright}$ and $t_{\triangleleft}$ and frequency extrema $f_{\triangleright}$ and $f_{\triangleleft}$ marked on figure 1 ought to provide relevant bounds for all five Morse eigenfunctions; the dotted lines in figure 2 show these extrema, and indeed they provide simultaneous bounds for all the first five eigenfunctions (with $f_{\triangleleft}$ seeming extremely conservative).

3. Multiple wavelet multivariate scalogram

(a) Vector-valued CWT

Let $\mathbf{X}(t) = [X_1(t), \dots, X_p(t)]^T$ denote the real p -vector-valued continuous-time time series consisting of a deterministic signal, $\mathbf{x}(t)$, plus random noise, $\boldsymbol{\eta}(t)$. Using the k th order generalized Morse analytic wavelet $\psi_k^+(t)$, (shorthand for $\psi_{k;\beta,\gamma}^+(t)$ for some choice of β and γ), we can define the column vector $\mathbf{W}_k^C(a, b; \mathbf{X}, \psi^+)$ to be the vector-valued version of the continuous wavelet transform (CWT):

$$\mathbf{W}_k^C(a, b; \mathbf{X}, \psi^+) = a^{-1/2} \int_{-\infty}^{\infty} \mathbf{X}(t) \psi_k^- \left(\frac{t-b}{a} \right) dt, \quad (3.1)$$

for $0 < a \leq \infty, -\infty \leq b \leq \infty$, where $\psi_k^-(\cdot)$ is the complex conjugate of $\psi_k^+(\cdot)$. Now $\mathbf{W}_k^C(a, b; \mathbf{X}, \psi^+) = a^{1/2} [\mathbf{X} * \psi_k'(b)]$, where $*$ denotes convolution and $\psi_k'(t) = \psi_k^-(-\frac{t}{a})/a$. The Fourier transform of $\psi_k'(t)$ is

$$\frac{1}{a} \int_{-\infty}^{\infty} \psi_k^- \left(-\frac{t}{a} \right) e^{-i2\pi ft} dt = [\Psi_k^+(af)]^* = \Psi_k^+(af), \quad (3.2)$$

since $\Psi_k^+(f)$ is real-valued. So the Fourier transform of the bracketed convolution is given by $\boldsymbol{\mathcal{X}}(f) \Psi_k^+(af)$, where $\boldsymbol{\mathcal{X}}(f)$ is the vector Fourier transform of $\mathbf{X}(t)$, and $\Psi_k^+(f)$ is the Fourier transform of $\psi_k^+(t)$. We can write $\mathbf{W}_k^C(a, b; \mathbf{X}, \psi^+)$ in terms of the inverse Fourier transform

$$\mathbf{W}_k^C(a, b; \mathbf{X}, \psi^+) = a^{1/2} \int_{-\infty}^{\infty} \boldsymbol{\mathcal{X}}(f) \Psi_k^+(af) e^{i2\pi fb} df. \quad (3.3)$$

(b) CWT via inverse DFT

Suppose we sample the series with a fine enough sample interval Δt to avoid aliasing, (i.e., $\boldsymbol{\mathcal{X}}(f) = 0$ for $|f| > 1/[2\Delta t]$) and use the shorthand $\mathbf{X}_t \equiv \mathbf{X}(t\Delta t)$, $t = 0, \dots, N-1$ and let $a = a_0\Delta t > 0$ and $b = b_0\Delta t$. Take N to be even. Then since $\Psi_k^+(f) = 0$ for $f < 0$, (3.3) can be discretized as

$$\begin{aligned} \mathbf{W}_k^C(a, b; \mathbf{X}, \psi^+) &= a^{1/2} \int_0^{\frac{1}{2\Delta t}} \boldsymbol{\mathcal{X}}(f) \Psi_k^+(af) e^{i2\pi fb} df \\ &\approx \frac{a^{1/2}}{N\Delta t} \sum_{l=0}^{N'-1} \boldsymbol{\mathcal{X}}\left(\frac{l}{N\Delta t}\right) \Psi_k^+\left(\frac{al}{N\Delta t}\right) e^{i2\pi lb/(N\Delta t)}, \end{aligned}$$

where $N' = N/2$. But $\boldsymbol{\mathcal{X}}(\frac{l}{N\Delta t}) \approx \boldsymbol{\mathcal{X}}_l = \Delta t \sum_{t=0}^{N-1} \mathbf{X}_t e^{-i2\pi tl/N}$, $l = 0, \dots, N-1$, the discrete Fourier transform (DFT) of $\{\mathbf{X}_t\}$, so we obtain a discretized form of the CWT which can be written in either of two equivalent forms:

$$\mathbf{W}_k(a, b; \mathbf{X}, \psi^+) = \frac{(a_0)^{1/2}}{N(\Delta t)^{1/2}} \sum_{l=0}^{N'-1} \boldsymbol{\mathcal{X}}_l \Psi_k^+\left(a_0 \frac{l}{N}\right) e^{i2\pi lb_0/N} \quad (3.4)$$

$$= \frac{(a_0)^{1/2}}{N(\Delta t)^{1/2}} \sum_{l=0}^{N-1} \boldsymbol{\mathcal{X}}_l \Psi_{k,l}^+(a_0) e^{i2\pi lb_0/N}, \quad (3.5)$$

Figure 3. Scalograms for the signal $x(t) = \psi_{0;8,3}^{(e)}(t)$. Left: Using the zeroth-order Morse wavelet, i.e., $|W_0(a, b; x)|^2$. Right: Using the first-order Morse wavelet, i.e., $|W_1(a, b; x)|^2$. The wavelet parameters used were $\beta = 8, \gamma = 3$.

where

$$\Psi_{k,l}^+(a_0) = \begin{cases} \Psi_k^+(a_0 \frac{l}{N}), & \text{for } l = 0, \dots, (N/2) - 1; \\ 0, & \text{for } l = (N/2), \dots, N - 1. \end{cases} \quad (3.6)$$

The discrete form $\mathbf{W}_k(a, b; \mathbf{X}, \psi^+)$ in (3.5) is an efficient formula for computation as it is a multiple of the inverse DFT of $\{\mathcal{X}_l \Psi_{k,l}^+(a_0), l = 0, \dots, N - 1\}$ and can be calculated using the inverse fast Fourier transform. The form $\mathbf{W}_k(a, b; \mathbf{X}, \psi^+)$ in (3.4) is somewhat more convenient for theoretical considerations. Hereafter, where use of the analytic wavelet ψ_k^+ is clearly understood, we shall use the briefer form $\mathbf{W}_k(a, b; \mathbf{X})$.

(c) Scalograms and energy concentration

Having now defined the CWT, we can at this point examine how the Morse wavelets concentrate signal in time and scale. Here we only use the scalar version of the CWT ($p = 1$) and consider analysis of the deterministic noise-free real-valued signal $x(t) = \psi_{0;8,3}^{(e)}(t)$, i.e., the even part of the zeroth-order Morse wavelet — see (2.3). Figure 3 shows the scalograms $|W_0(a, b; x)|^2$ and $|W_1(a, b; x)|^2$, (using Morse wavelets with $\beta = 8, \gamma = 3$) illustrating the energy concentration of the signal for the wavelets of orders zero and one. Because of the necessary orthogonality of the wavelets (eigenfunctions), there is for the first-order scalogram a ‘hole’ at the point where the zeroth-order scalogram shows highest energy concentration; the higher order also leads to more dissipated energy concentration. An analogous pattern extends to higher orders.

(d) *Multiple wavelet scalogram*

Using the vector $\{\mathbf{X}_t\}$ we can define a multiple wavelet (time-varying) scalogram estimate as the $p \times p$ matrix

$$\hat{\mathbf{S}}(a, b; \mathbf{X}) = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{W}_k(a, b; \mathbf{X}) \mathbf{W}_k^H(a, b; \mathbf{X}), \quad (3.7)$$

where H represents the Hermitian transpose.

In order to look at polarization as a function of time via parameter b , and scale via parameter a we need to look for energy concentration in linear combinations of the CWTs of the vector components of the signal. As in Park *et al* (1987) and Lilly and Park (1995) we do this using the singular value decomposition method.

4. Multiple wavelet polarization

(a) *Singular value decomposition*

Suppose we define the $K \times p$ matrix

$$\mathbf{M}(a, b; \mathbf{X}) = K^{-1/2} \begin{bmatrix} \mathbf{W}_0^H(a, b; \mathbf{X}) \\ \mathbf{W}_1^H(a, b; \mathbf{X}) \\ \vdots \\ \mathbf{W}_{K-1}^H(a, b; \mathbf{X}) \end{bmatrix}, \quad (4.1)$$

so that the l th column of $\mathbf{M}(a, b; \mathbf{X})$ contains the (conjugated) multiple wavelet transforms of the l th series. This matrix has rank $r_{\mathbf{M}} \leq \min(K, p)$. Actually, since we have a noisy system, we expect $r_{\mathbf{M}} = \min(K, p)$. Additionally, in what follows, we take $\min(K, p) = p$ so that $r_{\mathbf{M}} = p$.

For notational simplicity let us here drop explicit dependence on $(a, b; \mathbf{X})$. The singular value decomposition of $\mathbf{M}$ gives $\mathbf{M} = \mathbf{U} \mathbf{\Xi} \mathbf{V}^H$. (e.g., Strang 1988, p. 443). The columns $\{\mathbf{u}_l\}$ of the $K \times K$ unitary matrix $\mathbf{U}$ are eigenvectors of $\mathbf{M} \mathbf{M}^H$, the columns $\{\mathbf{v}_l\}$ of the $p \times p$ unitary matrix $\mathbf{V}$ are eigenvectors of $\mathbf{M}^H \mathbf{M}$, and the $r_{\mathbf{M}} = p$ singular values $\{\xi_l\}$ on the diagonal of the diagonal $K \times p$ matrix $\mathbf{\Xi}$ given by

$$\mathbf{\Xi} = \begin{bmatrix} \xi_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \xi_p \\ & \mathbf{0}_{K-p \times p} & & \end{bmatrix}$$

are the square roots of the nonzero eigenvalues of both $\mathbf{M} \mathbf{M}^H$ and $\mathbf{M}^H \mathbf{M}$. In other words,

$$\begin{aligned} \mathbf{M} \mathbf{v}_j &= \xi_j \mathbf{u}_j & \text{and} & & \mathbf{M}^H \mathbf{u}_i &= \xi_i \mathbf{v}_i \\ \mathbf{M}^H \mathbf{M} \mathbf{v}_j &= \xi_j^2 \mathbf{v}_j, j = 1, \dots, p & \text{and} & & \mathbf{M} \mathbf{M}^H \mathbf{u}_i &= \xi_i^2 \mathbf{u}_i, i = 1, \dots, K. \end{aligned} \quad (4.2)$$

With the singular values ranked in decreasing order of magnitude we have $\xi_i = \xi_j$, $i = j$, $i \leq r_{\mathbf{M}}$ and $\xi_i = 0$, $i > r_{\mathbf{M}}$ and $\xi_j = 0$, $j > r_{\mathbf{M}}$.

Then $\hat{\mathbf{S}}(a, b; \mathbf{X})$ in (3.7) can be written in the matrix form $\hat{\mathbf{S}} = \mathbf{M}^H \mathbf{M}$. The rotation of $\mathbf{M}$ by $\mathbf{V}$ gives, (with $r_{\mathbf{M}} = p$), $\mathbf{M} \mathbf{V} = \mathbf{U} \boldsymbol{\Xi} = [\xi_1 \mathbf{u}_1, \dots, \xi_p \mathbf{u}_p]$. If $\xi_1 \gg \xi_2, \dots, \xi_p$ nearly all the energy is explained by motion along a single element of the rotated basis set, and the components of $\mathbf{v}_1$, (often called the first right-singular vector of $\mathbf{M}$), are direction cosines of this motion (relative to the components of the acquisition instruments). So the polarization vector is the eigenvector $\mathbf{v}_1$ corresponding to the largest singular value ξ_1 (Park *et al*, 1987) and gives the direction of motion at scale/time (a, b) which contains the largest fraction of energy.

Now from (4.2) $\{\xi_l^2, l = 1, \dots, p\}$, and $\{\mathbf{v}_l, l = 1, \dots, p\}$ are the eigenvalues and eigenvectors of $\mathbf{M}^H \mathbf{M}$ and since the trace of a square matrix is the sum of its eigenvalues we get $\sum_{l=1}^p (\hat{\mathbf{S}})_{ll} = \sum_{l=1}^p \xi_l^2$, a sort of “preservation of energy” condition. Lilly and Park (1995) concentrate attention on the normalised singular values $\tilde{\xi}_l = \xi_l / \sqrt{\sum_{k=1}^p \xi_k^2}$ so that $\tilde{\xi}_l^2$ gives the proportion of energy explained by the polarization direction $\mathbf{v}_l$.

(b) Complex wavelets and phase of polarization components

If real wavelets were to be used in the above development then the principal polarization $\mathbf{v}_1$ would be real-valued and represent only rectilinear motion. In order to detect coherent motion with elliptical polarization, complex-valued wavelets are required, in order to have sensitivity to phase shifts among components. As we have noted, Lilly and Park (1995) introduced the approximately analytic multiple Slepian pseudo-wavelets, while here we concentrate on the exactly analytic multiple Morse wavelets; these classes are compared in section 6(c).

Now

$$\mathbf{M}^H \mathbf{M} = (\mathbf{V} \boldsymbol{\Xi} \mathbf{U}^H)(\mathbf{U} \boldsymbol{\Xi} \mathbf{V}^H) = \mathbf{V} \boldsymbol{\Xi} \mathbf{I}_r \boldsymbol{\Xi} \mathbf{V}^H = \mathbf{V} \boldsymbol{\Xi}^2 \mathbf{V}^H = \sum_{l=1}^p \xi_l^2 \mathbf{v}_l \mathbf{v}_l^H,$$

so if ξ_1 is the dominant singular value, $\mathbf{M}^H \mathbf{M} \approx \xi_1^2 \mathbf{v}_1 \mathbf{v}_1^H$, and

$$(\mathbf{M}^H \mathbf{M})_{lm} = \frac{1}{K} \sum_{k=0}^{K-1} (\mathbf{W}_k)_l (\mathbf{W}_k^H)_m \approx \xi_1^2 (\mathbf{v}_1)_l (\mathbf{v}_1^H)_m.$$

Putting $\hat{\mathbf{z}} = \mathbf{v}_1$ to emphasize the stochastic nature of this vector through $\mathbf{X}$, and writing the l th element $(\mathbf{W}_k)_l$ of $\mathbf{W}_k$ as $W_{k;l}$, and similarly for $\hat{\mathbf{z}}$, and re-introducing the dependence on $(a, b; \mathbf{X})$ for clarity, we obtain

$$\frac{\sum_k W_{k;l}(a, b; \mathbf{X}) W_{k;m}^H(a, b; \mathbf{X})}{[\sum_k |W_{k;l}(a, b; \mathbf{X})|^2 \sum_k |W_{k;m}(a, b; \mathbf{X})|^2]^{1/2}} \approx \frac{\hat{z}_l(a, b) \hat{z}_m^H(a, b)}{|\hat{z}_l(a, b)| |\hat{z}_m^H(a, b)|} = e^{i \hat{\mu}_{lm}(a, b)}, \quad (4.3)$$

where $\hat{\mu}_{lm}(a, b)$ is the phase difference between components l and m of $\hat{\mathbf{z}}(a, b)$.

5. The statistical structure using Morse wavelets

(a) Basics

Let $\mathbf{X}_t = \mathbf{x}_t + \boldsymbol{\eta}_t$, where $\{\mathbf{x}_t\}$ is a deterministic signal and $\{\boldsymbol{\eta}_t\}$ is a real-valued p -component multivariate Gaussian white noise with mean vector $E\{\boldsymbol{\eta}_t\} = \boldsymbol{\mu}$ and

covariance matrix $\text{cov}\{\boldsymbol{\eta}_t, \boldsymbol{\eta}_u\} = \sigma_\eta^2 \mathbf{I}_p \delta_{t,u} / \Delta t$, where $E\{\cdot\}$ denotes expectation, $\mathbf{I}_p$ denotes a p -dimensional identity matrix, $\delta_{t,u}$ denotes the Kronecker delta function; this covariance matrix implies that all three components have the same noise variance (isotropic) and the noise is uncorrelated between components.

We now give the resulting statistical structure of the $\mathbf{W}_k(a, b; \boldsymbol{\eta})$. A result made use of is that

$$\sum_{t=0}^{N-1} e^{\pm i 2\pi t k / N} = \begin{cases} N & \text{if } k = mN \text{ for integer } m; \\ 0, & \text{otherwise.} \end{cases} \quad (5.1)$$

(b) *Mean*

Using (3.4) the vector mean of the CWT of the noise $\{\boldsymbol{\eta}_t\}$ is given by

$$\begin{aligned} E\{\mathbf{W}_k(a, b; \boldsymbol{\eta})\} &= E\left\{ \frac{a^{1/2}}{N} \sum_{l=0}^{N'-1} \left(\sum_{t=0}^{N-1} \boldsymbol{\eta}_t e^{-i 2\pi t l / N} \right) \Psi_k^+(a_0 \frac{l}{N}) e^{i 2\pi l b_0 / N} \right\} \\ &= \frac{a^{1/2}}{N} \boldsymbol{\mu} \sum_{l=0}^{N'-1} N \delta_{l,0} \Psi_k^+(a_0 \frac{l}{N}) e^{i 2\pi l b_0 / N} = \boldsymbol{\mu} a^{1/2} \Psi_k^+(0) = \boldsymbol{\mu} a^{1/2} \int_{-\infty}^{\infty} \psi_k^+(t) dt = \mathbf{0}, \end{aligned}$$

since the wavelets are designed to integrate to zero. Hence the (vector) mean of the CWT of the noise is always zero, whatever is $E\{\boldsymbol{\eta}_t\} = \boldsymbol{\mu}$.

(c) *Covariance*

Now, since $E\{\mathbf{W}_k(a, b; \boldsymbol{\eta})\} = \mathbf{0}$, it follows that $\text{cov}\{\mathbf{W}_j(a, b; \boldsymbol{\eta}), \mathbf{W}_k(a, b; \boldsymbol{\eta})\} = E\{\mathbf{W}_j(a, b; \boldsymbol{\eta}) \mathbf{W}_k^H(a, b; \boldsymbol{\eta})\}$. It is shown in (A 1) of Appendix A that

$$E\{\mathbf{W}_j(a, b; \boldsymbol{\eta}) \mathbf{W}_k^H(a, b; \boldsymbol{\eta})\} = \frac{a_0 \sigma_\eta^2 \mathbf{I}_p}{N} \sum_{l=0}^{N'-1} \Psi_j^+(a_0 \frac{l}{N}) \Psi_k^+(a_0 \frac{l}{N}).$$

Approximation of this sum by an integral gives

$$\begin{aligned} E\{\mathbf{W}_j(a, b; \boldsymbol{\eta}) \mathbf{W}_k^H(a, b; \boldsymbol{\eta})\} &\approx \frac{a_0 \sigma_\eta^2 \mathbf{I}_p}{N} \int_0^{N'} \Psi_j^+(a_0 \frac{\nu}{N}) \Psi_k^+(a_0 \frac{\nu}{N}) d\nu \\ &= \sigma_\eta^2 \mathbf{I}_p \int_0^{\frac{a}{2\Delta t}} \Psi_j^+(f) \Psi_k^+(f) df = \sigma_\eta^2 \mathbf{I}_p \delta_{j,k}, \end{aligned}$$

as long as $a/(2\Delta t)$ exceeds f' , the frequency such that $\Psi_{K-1}^+(f) \approx 0$ for $f > f'$ (the ‘cut-off’ frequency of each wavelet in the frequency domain increases with order k — see Figure 2 where we could take f' considerably less than the conservative bound f_Δ).

Hence, since $E\{\mathbf{W}_k(a, b; \boldsymbol{\eta})\} = \mathbf{0}$, we have that (up to the discrete/continuous approximation error), $\text{cov}\{\mathbf{W}_j(a, b; \boldsymbol{\eta}), \mathbf{W}_k(a, b; \boldsymbol{\eta})\} = \sigma_\eta^2 \mathbf{I}_p \delta_{j,k}$.

In deriving the full suite of properties of these complex random variables, it is also important to evaluate $E\{\mathbf{W}_j(a, b; \boldsymbol{\eta}) \mathbf{W}_k^T(a, b; \boldsymbol{\eta})\}$. This is sometimes called the relation matrix and is needed in Appendix B. It is shown in (A 2) of Appendix A that $E\{\mathbf{W}_j(a, b; \boldsymbol{\eta}) \mathbf{W}_k^T(a, b; \boldsymbol{\eta})\} = \mathbf{0}$ for all j, k , which characterizes second-order circularity. The proof makes use of the fact that the Morse wavelets are supported on positive frequencies only.

(d) Real and imaginary parts

Since $\Psi^+(\cdot)$ is real,

$$\Re\{\mathbf{W}_k(a, b; \boldsymbol{\eta})\} = \frac{a^{1/2}}{N} \sum_{l=0}^{N'-1} \sum_{t=0}^{N-1} \boldsymbol{\eta}_t \Psi_k^+(a_0 \frac{l}{N}) \cos \left[\frac{2\pi l(b_0-t)}{N} \right],$$

with the imaginary part defined analogously. Now $E\{\mathbf{W}_k(a, b; \boldsymbol{\eta})\} = \mathbf{0}$, so the real and imaginary parts both have mean zero. The covariance of the real and imaginary parts then reduces to $E\{\Re\{\mathbf{W}_j(a, b; \boldsymbol{\eta})\} \Im\{\mathbf{W}_k^T(a, b; \boldsymbol{\eta})\}\}$ which takes the form

$$\begin{aligned} & \frac{a_0 \sigma_\eta^2 \mathbf{I}_p}{N^2} \sum_{l=0}^{N'-1} \sum_{m=0}^{N'-1} \Psi_j^+(a_0 \frac{l}{N}) \Psi_k^+(a_0 \frac{m}{N}) \sum_{t=0}^{N-1} \cos \left[\frac{2\pi l(b_0-t)}{N} \right] \sin \left[\frac{2\pi m(b_0-t)}{N} \right] \\ & + \frac{a \boldsymbol{\mu} \boldsymbol{\mu}^T}{N^2} \sum_{l=0}^{N'-1} \sum_{m=0}^{N'-1} \Psi_j^+(a_0 \frac{l}{N}) \Psi_k^+(a_0 \frac{m}{N}) \sum_{t=0}^{N-1} \cos \left[\frac{2\pi l(b_0-t)}{N} \right] \sum_{u=0}^{N-1} \sin \left[\frac{2\pi m(b_0-u)}{N} \right] = \mathbf{0}, \end{aligned}$$

because the sum of cosine/sine products is just a phase-shifted version of a well-known zero identity, while the sum of cosines equals $N \cos(2\pi l b_0/N) \delta_{l,0}$ and the sum of sines equals $N \sin(2\pi m b_0/N) \delta_{m,0} = 0$. Since the real and imaginary parts are uncorrelated and multivariate Gaussian, they are independent. The fact that $E\{\mathbf{W}_j(a, b; \boldsymbol{\eta}) \mathbf{W}_j^T(a, b; \boldsymbol{\eta})\} = \mathbf{0}$ implies that for each component (which we know has mean zero) the variance of the real part minus the variance of the imaginary part equals zero. Hence, the real and imaginary parts of the l th component $W_{k;l}(a, b; \boldsymbol{\eta})$ have the same semi-variance, $\sigma_\eta^2/2$.

(e) Distributional structure

Let us here denote $\mathbf{W}_k(a, b; \boldsymbol{\eta})$ by simply $\mathbf{W}_k$ to further reduce notation. Then the pK vector-valued random variable

$$\widetilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_0 \\ \vdots \\ \mathbf{W}_{K-1} \end{bmatrix} \stackrel{d}{=} N_{pK}^C(\mathbf{0}_{pK}, \sigma_\eta^2 \mathbf{I}_{pK}), \quad (5.2)$$

where $\stackrel{d}{=}$ denotes ‘equal in distribution,’ $\mathbf{0}_n$ denotes an n -vector of zeros, and $N_n^C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes an n vector-valued complex multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\widetilde{\mathbf{W}}$ satisfies (e.g., Brillinger 1981, p. 89) $E\{\widetilde{\mathbf{W}}\} = \mathbf{0}_{pK}$,

$$E\{(\widetilde{\mathbf{W}} - E\{\widetilde{\mathbf{W}}\})(\widetilde{\mathbf{W}} - E\{\widetilde{\mathbf{W}}\})^H\} = E\{\widetilde{\mathbf{W}} \widetilde{\mathbf{W}}^H\} = \sigma_\eta^2 \mathbf{I}_{pK} \text{ and} \quad (5.3)$$

$$E\{(\widetilde{\mathbf{W}} - E\{\widetilde{\mathbf{W}}\})(\widetilde{\mathbf{W}} - E\{\widetilde{\mathbf{W}}\})^T\} = E\{\widetilde{\mathbf{W}} \widetilde{\mathbf{W}}^T\} = \mathbf{0}_{pK}, \quad (5.4)$$

results which were shown in (b) and (c). This distribution follows since the $\{\boldsymbol{\eta}_t\}$ are jointly Gaussian and thus the linear combinations $\Re\{\mathbf{W}_0\}$, $\Im\{\mathbf{W}_0\}$, $\Re\{\mathbf{W}_1\}$, $\Im\{\mathbf{W}_1\}, \dots, \Re\{\mathbf{W}_{K-1}\}$ and $\Im\{\mathbf{W}_{K-1}\}$ are jointly Gaussian and, additionally us-

ing (d) above,

$$\begin{bmatrix} \Re\{\mathbf{W}_0\} \\ \Im\{\mathbf{W}_0\} \\ \vdots \\ \Re\{\mathbf{W}_{K-1}\} \\ \Im\{\mathbf{W}_{K-1}\} \end{bmatrix} \stackrel{d}{=} N_{2pK}(\mathbf{0}_{2pK}, \frac{1}{2}\sigma_\eta^2 \mathbf{I}_{2pK}), \quad (5.5)$$

where $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes an n vector-valued real multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; then (5.5) defines $\widetilde{\mathbf{W}} \stackrel{d}{=} N_{pK}^C(\mathbf{0}_{pK}, \sigma_\eta^2 \mathbf{I}_{pK})$ (Brillinger 1981, p. 89), as stated.

Note that the $\{W_{k;l}, k = 0, \dots, K-1; l = 1, \dots, p\}$, the components of $\widetilde{\mathbf{W}}$ are independent. Also, of course $\Re\{\widetilde{\mathbf{W}}\}$ and $\Im\{\widetilde{\mathbf{W}}\}$ are independent.

6. The phase of polarization components

(a) Phase uncertainty

From (4.3) we know that

$$\hat{\mu}_{lm}(a, b) = \tan^{-1} \left(\frac{\Im\{\hat{z}_l(a, b) \hat{z}_m^H(a, b)\}}{\Re\{\hat{z}_l(a, b) \hat{z}_m^H(a, b)\}} \right). \quad (6.1)$$

Dropping explicit dependence on $(a, b; \mathbf{X})$ temporarily, let

$$\hat{z}_l = z_l + \delta z_l \text{ and } \hat{z}_m = z_m + \delta z_m, \quad l \neq m \in \{1, \dots, p\},$$

where $z_l = E\{\hat{z}_l\}$ and $z_m = E\{\hat{z}_m\}$. Equivalently, $E\{\delta z_l\} = 0 = E\{\delta z_m\}$. In terms of real and complex components, let

$$\begin{aligned} \hat{z}_l &= u_1 + i u_2 \text{ and } \hat{z}_m = u_3 + i u_4 \\ z_l &= c_1 + i c_2 \text{ and } z_m = c_3 + i c_4 \\ \delta z_l &= \delta u_1 + i \delta u_2 \text{ and } \delta z_m = \delta u_3 + i \delta u_4. \end{aligned} \quad (6.2)$$

Hence,

$$\hat{\mu}_{lm} = \tan^{-1} \left(\frac{\Im\{\hat{z}_l \hat{z}_m^H\}}{\Re\{\hat{z}_l \hat{z}_m^H\}} \right) = \tan^{-1} \left(\frac{u_2 u_3 - u_1 u_4}{u_1 u_3 + u_2 u_4} \right).$$

First-order truncation of a Taylor series expansion (assuming $\|\delta \mathbf{z}\| < 1$) gives $\hat{\mu}_{lm} = \mu_{lm} + \delta \mu_{lm}$ where

$$\mu_{lm} = \hat{\mu}_{lm}|_{c_1, \dots, c_4} \text{ and } \delta \mu_{lm} = \sum_{j=1}^4 \delta u_j \frac{\partial \hat{\mu}_{lm}}{\partial u_j} \Big|_{c_1, \dots, c_4}.$$

Now

$$\frac{\partial \tan^{-1} \left(\frac{\Im\{w\}}{\Re\{w\}} \right)}{\partial u} = \frac{(\Re\{w\} \frac{\partial}{\partial u} \Im\{w\} - \Im\{w\} \frac{\partial}{\partial u} \Re\{w\})}{\Re^2\{w\} + \Im^2\{w\}}.$$

But since $\Re^2\{w\} + \Im^2\{w\} = |w|^2$, we obtain the real-valued quantity

$$\delta\mu_{lm} = \frac{|z_m|^2 \Im\{z_l^H \delta z_l\} - |z_l|^2 \Im\{z_m^H \delta z_m\}}{|z_l z_m|^2} = \Im\left\{ \frac{z_l^H \delta z_l}{|z_l|^2} - \frac{z_m^H \delta z_m}{|z_m|^2} \right\}.$$

Since $E\{\delta z_l\} = 0 = E\{\delta z_m\}$ we have that $E\{\delta\mu_{lm}\} = 0$, and so $\text{var}\{\delta\mu_{lm}\} = E\{\delta\mu_{lm}^2\}$.

With a_1, b_1, a_2, b_2 real-valued constants, we can write

$$Z = \left\{ \frac{z_l^H \delta z_l}{|z_l|^2} - \frac{z_m^H \delta z_m}{|z_m|^2} \right\} = (a_1 + ib_1)\delta z_l - (a_2 + ib_2)\delta z_m.$$

Let us look at the variance of the real and imaginary parts of Z . We note firstly that $E\{\delta \mathbf{z}\} = \mathbf{0}$. From (6.2), let $\delta z_l = \delta u_1 + i\delta u_2$ and $\delta z_m = \delta u_3 + i\delta u_4$. Then from (B15) of Appendix B, by putting $l = m = 1$, we have, to $O(\sigma_\eta^3)$, $E\{\delta \mathbf{z} \delta \mathbf{z}^T\} = \mathbf{0}$, which implies that $E\{\delta z_l \delta z_m\} = 0$ for all $l, m \in \{1, \dots, p\}$, and consequently, considering the cases $l \neq m$ and $l = m$,

$$\begin{aligned} E\{\delta u_1 \delta u_3\} &= E\{\delta u_2 \delta u_4\}; E\{\delta u_2 \delta u_3\} = -E\{\delta u_1 \delta u_4\}; E\{\delta u_1^2\} = E\{\delta u_2^2\} \\ E\{\delta u_3^2\} &= E\{\delta u_4^2\}, \text{ and } E\{\delta u_1 \delta u_2\} = E\{\delta u_3 \delta u_4\} = 0. \end{aligned}$$

Using these expectation equivalences we readily find that $\text{var}\{\Re\{Z\}\} = \text{var}\{\Im\{Z\}\}$.

Hence,

$$\begin{aligned} \text{var}\{\delta\mu_{lm}\} &= \frac{1}{2} E \left\{ \Re^2 \left\{ \frac{z_l^H \delta z_l}{|z_l|^2} - \frac{z_m^H \delta z_m}{|z_m|^2} \right\} + \Im^2 \left\{ \frac{z_l^H \delta z_l}{|z_l|^2} - \frac{z_m^H \delta z_m}{|z_m|^2} \right\} \right\} \\ &= \frac{1}{2} E \left\{ \left(\frac{z_l^H \delta z_l}{|z_l|^2} - \frac{z_m^H \delta z_m}{|z_m|^2} \right) \left(\frac{z_l \delta z_l^H}{|z_l|^2} - \frac{z_m \delta z_m^H}{|z_m|^2} \right) \right\} \\ &= \frac{1}{2} E \left\{ \frac{|\delta z_l|^2}{|z_l|^2} + \frac{|\delta z_m|^2}{|z_m|^2} - \frac{2\Re\{z_l z_m^H \delta z_m \delta z_l^H\}}{|z_l|^2 |z_m|^2} \right\} \\ &= \frac{1}{2} \left\{ \frac{\mathcal{V}_{ll}}{|z_l|^2} + \frac{\mathcal{V}_{mm}}{|z_m|^2} - \frac{2\Re\{z_l z_m^H \mathcal{V}_{ml}\}}{|z_l|^2 |z_m|^2} \right\}, \end{aligned} \quad (6.3)$$

where

$$\mathcal{V}_{ml} = E\{\delta z_m \delta z_l^H\} = (\text{cov}\{\delta \mathbf{z}, \delta \mathbf{z}\})_{ml} = (\text{cov}\{\delta \mathbf{v}_1, \delta \mathbf{v}_1\})_{ml}.$$

For three-component data ($p = 3$) this takes the form (see (B14), Appendix B)

$$\mathcal{V}_{ml} = \sigma_\eta^2 \frac{1}{K} \left(\frac{(\zeta_2 + \zeta_1)}{(\zeta_2 - \zeta_1)^2} \boldsymbol{\nu}_2 \boldsymbol{\nu}_2^H + \frac{(\zeta_3 + \zeta_1)}{(\zeta_3 - \zeta_1)^2} \boldsymbol{\nu}_3 \boldsymbol{\nu}_3^H \right)_{ml} + O(\sigma_\eta^3),$$

where ζ_1, ζ_2 and ζ_3 are the eigenvalues, in decreasing magnitude, of the noise-free scalogram $\mathbf{S}^{(0)} = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\mathbf{x})$, and $\boldsymbol{\nu}_2$ and $\boldsymbol{\nu}_3$ are the corresponding second and third eigenvectors.

The expression (6.3) represents a modified form of equation (A4) of Park *et al* (1987); the latter contains denominator terms affected by observational error.

Since here the first eigenvector relates to ‘signal’ and the latter two to ‘noise’ and $\lambda_1 \gg \lambda_2, \lambda_3$ with $\lambda_2, \lambda_3 \approx 0$, we have from (B 27) of Appendix B that an estimator for the covariance is given by

$$\hat{\mathcal{V}}_{ml} = (\widehat{\text{cov}}\{\delta\mathbf{v}_1, \delta\mathbf{v}_1\})_{ml} \approx \left(\frac{\lambda_2 + \lambda_3}{2(K-1)\lambda_1} [\mathbf{v}_2 \mathbf{v}_2^H + \mathbf{v}_3 \mathbf{v}_3^H] \right)_{ml}, \quad (6.4)$$

where λ_1, λ_2 and λ_3 are the eigenvalues of the noisy scalogram $\hat{\mathbf{S}} = \mathbf{M}^H \mathbf{M} = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{W}_k(\mathbf{X}) \mathbf{W}_k^H(\mathbf{X})$, in decreasing magnitude, and $\mathbf{v}_2$ and $\mathbf{v}_3$ are the corresponding second and third eigenvectors.

From (6.3) we thus obtain an *estimator* for $\text{var}\{\delta\mu_{lm}(a, b)\}$ given by

$$\widehat{\text{var}}\{\delta\mu_{lm}(a, b)\} = \frac{1}{2} \left\{ \frac{\hat{\mathcal{V}}_{ll}}{|\hat{z}_l|^2} + \frac{\hat{\mathcal{V}}_{mm}}{|\hat{z}_m|^2} - \frac{2\Re\{\hat{z}_l \hat{z}_m^H \hat{\mathcal{V}}_{ml}\}}{|\hat{z}_l|^2 |\hat{z}_m|^2} \right\},$$

where $\hat{\mathcal{V}}_{ml}$ is given by (6.4), and $\hat{z}_l = (\mathbf{v}_1)_l$, the l th component of the principal polarization vector.

(b) *Relationship of polarization phase to time series phase*

Any signal can be written as a sum of an analytic and an anti-analytic signal (e.g., Poletti, 1997). Let $x(t)$ be a real-valued scalar deterministic signal, zero outside the possibly infinite interval $[t_0, t_1]$. We consider it as a component of $\mathbf{x}(t)$. Then we can write $x(t)$ as:

$$x(t) = \frac{1}{2}[x(t) + i\mathcal{H}\{x(t)\}] + \frac{1}{2}[x(t) - i\mathcal{H}\{x(t)\}] = \frac{1}{2}x^+(t) + \frac{1}{2}x^-(t), \quad (6.5)$$

where $x^+(t) = x(t) + i\mathcal{H}\{x(t)\}$ is the analytic signal, and $x^-(t)$ the anti-analytic signal. If we let $\mathcal{F}\{\cdot\}$ denote the Fourier transform, so that $X(f) = \mathcal{F}\{x(t)\}$, then correspondence in the frequency domain is given by

$$X(f) = \frac{X(f)}{2}[1 + \text{sgn}(f)] + \frac{X(f)}{2}[1 - \text{sgn}(f)] \quad (6.6)$$

where $\text{sgn}(f) = 1$ if $f > 0$; 0 if $f = 0$, and -1 , if $f < 0$. The analytic and anti-analytic signals are complex-conjugates of each other. With $x^\pm(t) = R(t)e^{\pm i\varphi(t)}$, where $R(t)$ and $\varphi(t)$ are time-varying amplitude and phase functions,

$$\mathcal{F}\{x^\pm(t)\} = \frac{X(f)}{2}[1 \pm \text{sgn}(f)] = X^\pm(f), \text{ say.} \quad (6.7)$$

Further, $x(t) = \frac{1}{2}[R(t)e^{i\varphi(t)} + R(t)e^{-i\varphi(t)}] = R(t) \cos[\varphi(t)]$. Then from the definition of the continuous wavelet transform (3.1)

$$\begin{aligned} W^C(a, b; x, \psi^+) &= a^{-1/2} \int_{-\infty}^{\infty} R(t) \cos[\varphi(t)] \psi^-\left(\frac{t-b}{a}\right) dt \\ &= a^{-1/2} \int_{-\infty}^{\infty} \frac{1}{2} R(t) e^{i\varphi(t)} \psi^-\left(\frac{t-b}{a}\right) dt \end{aligned} \quad (6.8)$$

$$+ a^{-1/2} \int_{-\infty}^{\infty} \frac{1}{2} R(t) e^{-i\varphi(t)} \psi^-\left(\frac{t-b}{a}\right) dt. \quad (6.9)$$

Figure 4. Grey area shows total domain Ω comprising all scale-time points (a, b) , $a > 0$, at which the phase can be recovered; it comprises subdomains $\Omega_1, \dots, \Omega_4$.

But from (3.3), (6.6) and (6.7),

$$\begin{aligned}\mathcal{F}\{W^C(a, b; x, \psi^+)\} &= a^{1/2} X(f) \Psi^+(af) = a^{1/2} [X^+(f) + X^-(f)] \Psi^+(af) \\ &= a^{1/2} X^+(f) \Psi^+(af),\end{aligned}$$

since $x^-(t)$ is anti-analytic. Hence, using (6.7) inverse Fourier transformation gives

$$\begin{aligned}W^C(a, b; x, \psi^+) &= a^{1/2} \int_{-\infty}^{\infty} X^+(f) \Psi^+(af) e^{i2\pi bf} df \\ &= a^{1/2} \iint_{-\infty}^{\infty} \frac{1}{2} R(t) e^{i\varphi(t)} \Psi^+(af) e^{i2\pi(b-t)f} df dt.\end{aligned}$$

Then, using (3.2), $W^C(a, b; x, \psi^+) = a^{-1/2} \int_{-\infty}^{\infty} \frac{1}{2} R(t) e^{i\varphi(t)} \psi^-\left(\frac{t-b}{a}\right) dt$, so that a comparison with (6.8) and (6.9) reveals that

$$a^{-1/2} \int_{-\infty}^{\infty} \frac{1}{2} R(t) e^{-i\varphi(t)} \psi^-\left(\frac{t-b}{a}\right) dt = 0. \quad (6.10)$$

Now assume that $\psi^-(t) = 0$ for all t outside $[t_{\min}, t_{\max}]$; note that $t_{\min} < 0$ and $t_{\max} > 0$; with reference to Fig. 2 we could take $t_{\min} \approx t_{\triangleright}$ and $t_{\max} \approx t_{\triangleleft}$. Then, for $a > 0$, the compact support of $\psi^-(t)$ combined with a translation by b and scaling by a , means that $\psi^-\left(\frac{t-b}{a}\right) = 0$ for all t outside $[b + a t_{\min}, b + a t_{\max}]$.

Suppose we let

$$y(t) = \begin{cases} R_0(t) \cos[\varphi(t) + \theta], & \text{if } t_0 \leq t \leq t_1; \\ 0, & \text{otherwise,} \end{cases} \quad (6.11)$$

where $R_0(t) = hR(t)$ for some real positive constant h , i.e., $y(t)$ is a phase-shifted version of rescaled $x(t)$. Moreover, since this phase relationship holds only for $t_0 \leq$

$t \leq t_1$, it equates to a *transient* phase relationship. Then

$$W^C(a, b; y, \psi^+) = \frac{1}{a^{1/2}} \int_{-\infty}^{\infty} y(t) \psi^-\left(\frac{t-b}{a}\right) dt = \frac{h}{a^{1/2}} \int_{t_L}^{t_U} R(t) \cos[\varphi(t) + \theta] \psi^-\left(\frac{t-b}{a}\right) dt,$$

where $t_L = \max\{t_0, b + a t_{\min}\}$ and $t_U = \min\{t_1, b + a t_{\max}\}$, (reflecting the finite supports of $y(t)$ and $\psi^-([t-b]/a)$). So

$$W^C(a, b; y, \psi^+) = \frac{he^{i\theta}}{2a^{1/2}} \int_{t_L}^{t_U} R(t) e^{i\varphi(t)} \psi^-\left(\frac{t-b}{a}\right) dt + \frac{he^{-i\theta}}{2a^{1/2}} \int_{t_L}^{t_U} R(t) e^{-i\varphi(t)} \psi^-\left(\frac{t-b}{a}\right) dt.$$

But from (6.10), the second integral is in fact zero, so that

$$\begin{aligned} W^C(a, b; y, \psi^+) &= \frac{he^{i\theta}}{2a^{1/2}} \int_{t_L}^{t_U} R(t) e^{i\varphi(t)} \psi^-\left(\frac{t-b}{a}\right) dt = \frac{he^{i\theta}}{2a^{1/2}} \int_{-\infty}^{\infty} R(t) e^{i\varphi(t)} \psi^-\left(\frac{t-b}{a}\right) dt \\ &= he^{i\theta} W^C(a, b; x, \psi^+). \end{aligned}$$

So $W^C(a, b; y, \psi^+) \propto e^{i\theta} W^C(a, b; x, \psi^+)$ for $\Omega = \{(a, b), a > 0 : t_L < t_U\}$, i.e., a Morse wavelet can retrieve a phase shift θ .

So if we let $x_m(t) = x(t)$ and $x_l(t) = y(t)$, be two components of a three-component deterministic signal $\mathbf{x}$, then, using the k th order generalized Morse wavelet $\psi_k^+(t)$ we get $W_{k;l}(a, b) \propto e^{i\theta} W_{k;m}(a, b)$, for (a, b) in Ω , (ignoring any error in moving from the continuous wavelet transform W^C to its discretized form, W). Note that if several phase shifted components were present in the signal then due to the linearity of the wavelet transform we would recover the phase shift at that component which had large magnitude at the relevant (a, b) point we were looking at. We then find,

$$\frac{\sum_k W_{k;l}(a, b; \mathbf{x}) W_{k;m}^H(a, b; \mathbf{x})}{[\sum_k |W_{k;l}(a, b; \mathbf{x})|^2 \sum_k |W_{k;m}(a, b; \mathbf{x})|^2]^{1/2}} = e^{i\theta_{lm}}, \quad (6.12)$$

for (a, b) in Ω . With reference to section 5(a) note that since $E\{\delta\mu_{lm}\} = 0$, and $\text{var}\{\delta\mu_{lm}\} \propto \sigma_\eta^2$, it follows that when $\sigma_\eta^2 = 0$, we have $\hat{\mu}_{lm} = \mu_{lm}$, and since in this case $\mathbf{X} = \mathbf{x}$, we have also (via a comparison of (4.3) and (6.12)) $\hat{\mu}_{lm} = \theta_{lm}$; hence $\mu_{lm} = \theta_{lm}$, as expected. Thus $\hat{\mu}_{lm} = \theta_{lm} + \delta\mu_{lm}$, for (a, b) in Ω , i.e., we can approximate the phase θ_{lm} via $\hat{\mu}_{lm}$ (up to integer multiples of 2π), with the accuracy of the approximation depending on the noise level, manifested through $\delta\mu_{lm}$.

The domain Ω will consist of four subdomains:

$$\begin{aligned} \Omega_1 &= \{(a, b), a > 0 : b + at_{\min} \geq t_0 \text{ and } b + at_{\max} \leq t_1\} \\ \Omega_2 &= \{(a, b), a > 0 : b + at_{\min} \leq t_0 \text{ and } b + at_{\max} \geq t_0 \\ &\quad \text{and } b + at_{\max} \leq t_1\} \\ \Omega_3 &= \{(a, b), a > 0 : b + at_{\min} \geq t_0 \text{ and } b + at_{\min} \leq t_1 \\ &\quad \text{and } b + at_{\max} \geq t_1\} \\ \Omega_4 &= \{(a, b), a > 0 : b + at_{\min} \leq t_0 \text{ and } b + at_{\max} \geq t_1\}. \end{aligned}$$

These subdomains are illustrated in figure 4 for a finite interval of the time axis; Ω is given by the total grey area which is of the form of a wedge which widens with increasing scale $a > 0$.

Figure 5. Left column: Phase estimates (of $\pi/2$) for the first four Morse wavelets. Middle column: Phase estimates for the first four Slepian pseudo-wavelets. Right column: Modulus squared of the Fourier transform of the Slepian pseudo-wavelets illustrating non-analyticity (frequency support not purely on positive frequencies).

(c) *Morse wavelets versus Slepian pseudo-wavelets*

At this point we can usefully compare the performance of Morse wavelets and Slepian pseudo-wavelets for phase-recovery. We consider the case of two sinusoids x_t and y_t localised at $f = 0.07$ and shifted from one another by $\pi/2$. Hence we wish to estimate the phase shift of $\pi/2$. For illustration purposes we compare the first four wavelets of each class.

For the Slepian pseudo-wavelets $\tilde{\psi}_k^{(S)}(f_c), k = 0, \dots, 3$ defined in (1.1) we followed an example in Lilly and Park (1995) using 100 point wavelets with $f_c = 0.07$ and $f_w = 0.035$. The resulting wavelet transform of x_t using the k th pseudo-wavelet is here denoted $W_k(f_c, t; x, \tilde{\psi}^{(S)})$. For $k = 0, \dots, 3$ the phase shift was estimated as

$$\hat{\theta}_k^{(S)}(t) = \tan^{-1} \left\{ \Im \left[\frac{W_k(f_c, t; y, \tilde{\psi}^{(S)})}{W_k(f_c, t; x, \tilde{\psi}^{(S)})} \right] \middle/ \Re \left[\frac{W_k(f_c, t; y, \tilde{\psi}^{(S)})}{W_k(f_c, t; x, \tilde{\psi}^{(S)})} \right] \right\}.$$

For the Morse wavelets we used $\psi_{k;\beta,\gamma}^+(t), k = 0, \dots, 3$ with $\beta = 8, \gamma = 3$. We make the usual identification (e.g., Mallat, 1998, p. 82) between scale and frequency, $a = f_0^{(k)}/f$ where $f_0^{(k)}$ is the reference frequency for the k th-order wavelet. A complex exponential $e^{2i\pi f_1 t}$, $f_1 > 0$, has Wigner distribution $\delta(f - f_1)$, and k th-

order Morse scalogram $|a|[\Psi_{k;\beta,\gamma}^+(af_1)]^2$. The Wigner distribution has its maximum at frequency f_1 and the scalogram will have its maximum at a scale $a = f_0^{(k)}/f_1$. So $f_0^{(k)}$ should be chosen to maximise $(f_0^{(k)}/f_1)[\Psi_{k;\beta,\gamma}^+(f_0^{(k)})]^2$, i.e., we choose

$$f_0^{(k)} : f_0^{(k)} \text{ maximises } f \left[\Psi_{k;\beta,\gamma}^+(f) \right]^2.$$

With $\beta = 8, \gamma = 3$ we obtain $f_0^{(0)} = 0.2252, f_0^{(1)} = 0.2737, f_0^{(2)} = 0.3060$ and $f_0^{(3)} = 0.3319$. Hence we evaluated the multiple wavelet transform at scale $a_k = f_0^{(k)}/0.07$. For $k = 0, \dots, 3$ the phase shift was estimated via

$$\hat{\theta}_k(b) = \tan^{-1} \left\{ \Im \left[\frac{W_k(a_k, b; y, \psi^+)}{W_k(a_k, b; x, \psi^+)} \right] \middle/ \Re \left[\frac{W_k(a_k, b; y, \psi^+)}{W_k(a_k, b; x, \psi^+)} \right] \right\}.$$

Figure 5 shows in the left column the phase estimates (of $\pi/2$) for the first four Morse wavelets. The middle column shows the phase estimates for the first four Slepian pseudo-wavelets — the 0th and 2nd-order Slepian pseudo-wavelets do a reasonable job of phase estimation, but show clear oscillatory behaviour; the 1st and 3rd order wavelets produce wildly oscillating results. In all cases the Morse wavelets perform better. The Morse wavelets are *exactly* analytic, while the Slepian pseudo-wavelets are only approximately analytic. The modulus squared of the Fourier transform of the Slepian pseudo-wavelets is shown in the right column of figure 5. The negative frequency contributions result from the wavelets being only approximately analytic and are clearly very costly in terms of the phase estimation performance.

7. Summary

We have demonstrated that multiple generalized Morse wavelets have the necessary properties for the extraction of transient phase information from multiple component recordings. The exact analyticity of these wavelets gives them an advantage over the approximately analytic Slepian pseudo-wavelets. A detailed statistical analysis framework has been provided, including a careful study of approximation errors. Examples of the highly successful application of the methodology presented here to one synthetic and two real 3-component recordings are given in Part II of this paper (Olhede and Walden 2002b).

Sofia Olhede is Beit Scientific Research Fellow, recipient of an EPSRC (UK) grant, and beneficiary of the Helge Ax:son Johnsons Stiftelse. The authors would like to thank the referees for helpful comments leading to an improved exposition.

Appendix A.

(a) Wavelet transform joint expectations

Here we compute $E\{\mathbf{W}_j(a, b; \boldsymbol{\eta})\mathbf{W}_k^H(a, b; \boldsymbol{\eta})\}$; this is given by

$$\frac{a}{N^2} E \left\{ \sum_{l=0}^{N'-1} \left(\sum_{t=0}^{N-1} \boldsymbol{\eta}_t e^{-i2\pi tl/N} \right) \Psi_j^+ \left(a_0 \frac{l}{N} \right) e^{i2\pi lb_0/N} \times \right. \\ \left. \sum_{m=0}^{N'-1} \left(\sum_{u=0}^{N-1} \boldsymbol{\eta}_u^H e^{i2\pi um/N} \right) \Psi_k^+ \left(a_0 \frac{m}{N} \right) e^{-i2\pi mb_0/N} \right\}.$$

So, denoting this expectation more simply by $E\{\mathbf{W}_j \mathbf{W}_k^H\}$, we get

$$\begin{aligned} E\{\mathbf{W}_j \mathbf{W}_k^H\} &= \frac{a}{N^2} \sum_{l=0}^{N'-1} \sum_{m=0}^{N'-1} \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} E\{\boldsymbol{\eta}_t \boldsymbol{\eta}_u^H\} e^{-i2\pi(tl-um)/N} \times \\ &\quad \Psi_j^+ \left(a_0 \frac{l}{N} \right) \Psi_k^+ \left(a_0 \frac{m}{N} \right) e^{-i2\pi(m-l)b_0/N} \\ &= \frac{a}{N^2} \sum_{l=0}^{N'-1} \sum_{m=0}^{N'-1} \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} \left[\frac{\sigma_\eta^2 \mathbf{I}_p \delta_{t,u}}{\Delta t} + \boldsymbol{\mu} \boldsymbol{\mu}^H \right] e^{-i2\pi(tl-um)/N} \times \\ &\quad \Psi_j^+ \left(a_0 \frac{l}{N} \right) \Psi_k^+ \left(a_0 \frac{m}{N} \right) e^{-i2\pi(m-l)b_0/N}, \end{aligned}$$

i.e.,

$$\begin{aligned} E\{\mathbf{W}_j \mathbf{W}_k^H\} &= \frac{a}{N^2} \sum_{l=0}^{N'-1} \sum_{m=0}^{N'-1} \left\{ \frac{\sigma_\eta^2 \mathbf{I}_p}{\Delta t} \left[\sum_{t=0}^{N-1} e^{i2\pi t(m-l)/N} \right] + \boldsymbol{\mu} \boldsymbol{\mu}^H \left[\sum_{t=0}^{N-1} e^{-i2\pi tl/N} \right. \right. \\ &\quad \left. \left. \sum_{u=0}^{N-1} e^{i2\pi um/N} \right] \right\} \Psi_j^+ \left(a_0 \frac{l}{N} \right) \Psi_k^+ \left(a_0 \frac{m}{N} \right) e^{-i2\pi(m-l)b_0/N}. \end{aligned}$$

Then, using (5.1),

$$\begin{aligned} E\{\mathbf{W}_j \mathbf{W}_k^H\} &= \frac{a}{N^2} \sum_{l=0}^{N'-1} \sum_{m=0}^{N'-1} \left\{ \frac{N \sigma_\eta^2 \mathbf{I}_p \delta_{l,m}}{\Delta t} + \boldsymbol{\mu} \boldsymbol{\mu}^H N^2 \delta_{l,0} \delta_{m,0} \right\} \times \\ &\quad \Psi_j^+ \left(a_0 \frac{l}{N} \right) \Psi_k^+ \left(a_0 \frac{m}{N} \right) e^{-i2\pi(m-l)b_0/N} \\ &= \frac{a_0 \sigma_\eta^2 \mathbf{I}_p}{N} \sum_{l=0}^{N'-1} \Psi_j^+ \left(a_0 \frac{l}{N} \right) \Psi_k^+ \left(a_0 \frac{l}{N} \right), \end{aligned} \tag{A 1}$$

since $\Psi_j^+(0) = \Psi_k^+(0) = 0$. We also compute $E\{\mathbf{W}_j(a, b; \boldsymbol{\eta}, \psi^+) \mathbf{W}_k^T(a, b; \boldsymbol{\eta}, \psi^+)\}$. Denoting this expectation more simply by $E\{\mathbf{W}_j \mathbf{W}_k^T\}$,

$$\begin{aligned} E\{\mathbf{W}_j \mathbf{W}_k^T\} &= \frac{a}{N^2} \sum_{l=0}^{N'-1} \sum_{m=0}^{N'-1} \left\{ \frac{\sigma_\eta^2 \mathbf{I}_p}{\Delta t} \left[\sum_{t=0}^{N-1} e^{-i2\pi t(m+l)/N} \right] + \boldsymbol{\mu} \boldsymbol{\mu}^T \left[\sum_{t=0}^{N-1} e^{-i2\pi tl/N} \right. \right. \\ &\quad \left. \left. \sum_{u=0}^{N-1} e^{-i2\pi um/N} \right] \right\} \Psi_j^+ \left(a_0 \frac{l}{N} \right) \Psi_k^+ \left(a_0 \frac{m}{N} \right) e^{i2\pi(m+l)b_0/N}. \end{aligned}$$

Then, letting $n = -m$,

$$\begin{aligned}
E\{\mathbf{W}_j \mathbf{W}_k^T\} &= \frac{a}{N^2} \sum_{l=0}^{N'-1} \sum_{n=1-N'}^0 \frac{\sigma_\eta^2 \mathbf{I}_p}{\Delta t} \left[\sum_{t=0}^{N-1} e^{-i2\pi t(l-n)/N} \right] \times \\
&\quad \Psi_j^+ \left(a_0 \frac{l}{N}\right) \Psi_k^+ \left(-a_0 \frac{n}{N}\right) e^{-i2\pi(n-l)b_0/N} + a\boldsymbol{\mu}\boldsymbol{\mu}^T \Psi_j^+(0) \Psi_k^+(0) \\
&= \frac{a_0 \sigma_\eta^2 \mathbf{I}_p}{N} \sum_{l=0}^{N'-1} \sum_{n=1-N'}^0 \delta_{l,n} \Psi_j^+ \left(a_0 \frac{l}{N}\right) \Psi_k^+ \left(-a_0 \frac{n}{N}\right) e^{-i2\pi(n-l)b_0/N} \\
&= \frac{a_0 \sigma_\eta^2 \mathbf{I}_p}{N} \Psi_j^+(0) \Psi_k^+(0) = \mathbf{0}.
\end{aligned} \tag{A 2}$$

Appendix B.

(a) Second Order Properties of the Eigenvectors

Here we derive $\text{cov}\{\boldsymbol{\delta}\mathbf{v}_l, \boldsymbol{\delta}\mathbf{v}_m\}$, where $\boldsymbol{\delta}\mathbf{v}_l = \mathbf{v}_l - \boldsymbol{\nu}_l$, the difference between the eigenvector of the noisy matrix $\hat{\mathbf{S}}(\mathbf{X})$ and the eigenvector of the noise-free version of this matrix, $\mathbf{S}^{(0)}$, say. Our approach broadly follows that of Park and Chave (1984), which in turn draws on Backus *et al* (1981); we hope our formulation is clear without skimping on important details.

Now $\mathbf{X}_t = \mathbf{x}_t + \boldsymbol{\eta}_t = \mathbf{x}_t + \sigma_\eta \boldsymbol{\epsilon}_t$, say. Then $\mathbf{W}_k(\mathbf{X}) = \mathbf{W}_k(\mathbf{x}) + \sigma_\eta \mathbf{W}(\boldsymbol{\epsilon})$ and

$$\hat{\mathbf{S}}(\mathbf{X}) = \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{W}_k(\mathbf{X}) \mathbf{W}_k^H(\mathbf{X}) = \mathbf{S}^{(0)} + \sigma_\eta \mathbf{S}^{(1)} + \sigma_\eta^2 \mathbf{S}^{(2)}, \tag{B 1}$$

where

$$\begin{aligned}
\mathbf{S}^{(0)} &= \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\mathbf{x}); \quad \mathbf{S}^{(1)} = \frac{1}{K} \sum_{k=0}^{K-1} [\mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\boldsymbol{\epsilon}) + \mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_k^H(\mathbf{x})] \\
\text{and } \mathbf{S}^{(2)} &= \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_k^H(\boldsymbol{\epsilon}).
\end{aligned}$$

The matrices $\hat{\mathbf{S}}(\mathbf{X})$, $\mathbf{S}^{(0)}$, $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are each complex Hermitian and hence have linear elementary divisors so that there exist p eigenvectors (with corresponding real eigenvalues) spanning $\mathbb{C}^p$, (Wilkinson 1965, pp. 26, 42). Further, the Hermitian property guarantees that the eigenvalues and eigenvectors of these matrices may be expanded in convergent power series (Kato, 1966, pp. 120–1).

Let $\{\xi_l^2\}$ and $\{\bar{\mathbf{v}}_l\}$ be the l th eigenvalue and *unstandardized* eigenvector of $\hat{\mathbf{S}}(\mathbf{X})$. Putting $\lambda_l = \xi_l^2$ we have

$$\hat{\mathbf{S}}(\mathbf{X}) \bar{\mathbf{v}}_l = \lambda_l \bar{\mathbf{v}}_l, \quad l = 1, \dots, p. \tag{B 2}$$

We are really interested in the eigenvectors of $\mathbf{S}^{(0)}$, (involving the error-free data),

$$\mathbf{S}^{(0)} \boldsymbol{\nu}_l = \zeta_l \boldsymbol{\nu}_l, \tag{B 3}$$

say. Let us expand the eigenvector $\bar{\mathbf{v}}_l$ in a convergent power series:

$$\bar{\mathbf{v}}_l = \sum_{k=0}^{\infty} \sigma_{\eta}^k \mathbf{v}_l^{(k)}, \quad (\text{B } 4)$$

where $\mathbf{v}_l^{(0)} = \boldsymbol{\nu}_l$, the corresponding eigenvector of $\mathbf{S}^{(0)}$. We shall take $\boldsymbol{\nu}_l$ to be a unit-norm eigenvector, so that $\langle \boldsymbol{\nu}_l, \boldsymbol{\nu}_l \rangle = 1$. Then for the eigenvector $\bar{\mathbf{v}}_l$,

$$\begin{aligned} \langle \bar{\mathbf{v}}_l, \bar{\mathbf{v}}_l \rangle &= \left\langle \sum_{j=0}^{\infty} \sigma_{\eta}^j \mathbf{v}_l^{(j)}, \sum_{k=0}^{\infty} \sigma_{\eta}^k \mathbf{v}_l^{(k)} \right\rangle = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sigma_{\eta}^{j+k} \langle \mathbf{v}_l^{(j)}, \mathbf{v}_l^{(k)} \rangle \\ &= \langle \boldsymbol{\nu}_l, \boldsymbol{\nu}_l \rangle + 2\sigma_{\eta} \Re\{\langle \boldsymbol{\nu}_l, \mathbf{v}_l^{(1)} \rangle\} + O(\sigma_{\eta}^2), \end{aligned}$$

so if we wish $\langle \bar{\mathbf{v}}_l, \bar{\mathbf{v}}_l \rangle = 1 + O(\sigma_{\eta}^2)$, then we must have $\Re\{\langle \boldsymbol{\nu}_l, \mathbf{v}_l^{(1)} \rangle\} = \Re\{\boldsymbol{\nu}_l^H \mathbf{v}_l^{(1)}\} = 0$. A sufficient condition for this is that the two vectors are orthogonal,

$$\langle \boldsymbol{\nu}_l, \mathbf{v}_l^{(1)} \rangle = 0. \quad (\text{B } 5)$$

We can also expand the eigenvalues λ_l in convergent power series,

$$\lambda_l = \sum_{k=0}^{\infty} \sigma_{\eta}^k \lambda_l^{(k)}, \quad (\text{B } 6)$$

where $\lambda_l^{(0)} = \zeta_l$, the corresponding eigenvalue of $\mathbf{S}^{(0)}$. We do not consider the possibility that noise has changed the order of the eigenvalues. As pointed out in Park and Chave (1984, p. 687) this is likely to be a problem only where the difference between two eigenvalues of $\mathbf{S}^{(0)}$ is comparable to the noise variance.

Then (B 1), (B 2), (B 4) and (B 6) give

$$\left(\mathbf{S}^{(0)} + \sigma_{\eta} \mathbf{S}^{(1)} + \sigma_{\eta}^2 \mathbf{S}^{(2)} \right) \sum_{k=0}^{\infty} \sigma_{\eta}^k \mathbf{v}_l^{(k)} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sigma_{\eta}^{m+n} \lambda_l^{(m)} \mathbf{v}_l^{(n)}. \quad (\text{B } 7)$$

Equating $O(1)$ terms gives (B 3). Equating $O(\sigma_{\eta})$ terms gives

$$\mathbf{S}^{(0)} \mathbf{v}_l^{(1)} + \mathbf{S}^{(1)} \boldsymbol{\nu}_l = \zeta_l \mathbf{v}_l^{(1)} + \lambda_l^{(1)} \boldsymbol{\nu}_l, \quad (\text{B } 8)$$

$$\text{i.e., } (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p) \mathbf{v}_l^{(1)} = (\lambda_l^{(1)} \mathbf{I}_p - \mathbf{S}^{(1)}) \boldsymbol{\nu}_l. \quad (\text{B } 9)$$

Let us assume that $\zeta_l \geq 0$ is a non-degenerate eigenvalue, i.e., there is not more than one corresponding eigenvector. Since $\mathbf{S}^{(0)}$ has a complete orthonormal basis for $\mathbb{C}^p$ of eigenvectors $\{\boldsymbol{\nu}_k, k = 1, \dots, p\}$, $\mathbf{S}^{(0)} = \sum_{k=1}^p \zeta_k \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H$ and $\mathbf{I}_p = \sum_{k=1}^p \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H$. Hence,

$$\begin{aligned} (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p) &= \sum_{j \neq l} (\zeta_j - \zeta_l) \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H \\ \text{and } (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p)^+ &= \sum_{k \neq l} (\zeta_k - \zeta_l)^{-1} \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H, \end{aligned}$$

the expression for the generalized inverse holding since ζ_l is assumed non-degenerate so that $\zeta_k \neq \zeta_l$ for any $k \neq l$. Thus,

$$\begin{aligned} (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p)^+ (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p) &= \sum_{k \neq l} \sum_{j \neq l} \frac{(\zeta_j - \zeta_l)}{(\zeta_k - \zeta_l)} \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H \\ &= \sum_{k \neq l} \sum_{j \neq l} \frac{(\zeta_j - \zeta_l)}{(\zeta_k - \zeta_l)} \boldsymbol{\nu}_k \delta_{k,j} \boldsymbol{\nu}_j^H = \sum_{k \neq l} \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H = \mathbf{I}_p - \boldsymbol{\nu}_l \boldsymbol{\nu}_l^H. \end{aligned}$$

So multiplying (B 9) by $(\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p)^+$ we get,

$$\begin{aligned} (\mathbf{I}_p - \boldsymbol{\nu}_l \boldsymbol{\nu}_l^H) \mathbf{v}_l^{(1)} &= (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p)^+ (\lambda_l^{(1)} \mathbf{I}_p - \mathbf{S}^{(1)}) \boldsymbol{\nu}_l \\ \Rightarrow \mathbf{v}_l^{(1)} &= \boldsymbol{\nu}_l \boldsymbol{\nu}_l^H \mathbf{v}_l^{(1)} + \lambda_l^{(1)} (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p)^+ \boldsymbol{\nu}_l - (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p)^+ \mathbf{S}^{(1)} \boldsymbol{\nu}_l. \end{aligned}$$

But $(\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p)^+ \boldsymbol{\nu}_l = \sum_{k \neq l} (\zeta_k - \zeta_l) \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H \boldsymbol{\nu}_l = \sum_{k \neq l} (\zeta_k - \zeta_l) \boldsymbol{\nu}_k \delta_{k,l} = \mathbf{0}_p$, so

$$\mathbf{v}_l^{(1)} = \boldsymbol{\nu}_l \langle \boldsymbol{\nu}_l, \mathbf{v}_l^{(1)} \rangle - (\mathbf{S}^{(0)} - \zeta_l \mathbf{I}_p)^+ \mathbf{S}^{(1)} \boldsymbol{\nu}_l. \quad (\text{B } 10)$$

Recalling (B 5), we obtain the unique solution

$$\mathbf{v}_l^{(1)} = \frac{1}{K} \sum_{j \neq l} \sum_{k=0}^{K-1} (\zeta_l - \zeta_j)^{-1} \boldsymbol{\nu}_j \{ \boldsymbol{\nu}_j^H [\mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\boldsymbol{\epsilon}) + \mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_k^H(\mathbf{x})] \boldsymbol{\nu}_l \}, \quad (\text{B } 11)$$

where we point out that everything inside the curly bracket is scalar in form.

From (B 4), $\bar{\mathbf{v}}_l = \boldsymbol{\nu}_l + \sigma_\eta \mathbf{v}_l^{(1)} + O(\sigma_\eta^2)$, so to first order in σ_η we have that $E\{\bar{\mathbf{v}}_l\} = \boldsymbol{\nu}_l + \sigma_\eta E\{\mathbf{v}_l^{(1)}\} = \boldsymbol{\nu}_l$, since $E\{\mathbf{W}_k^H(\boldsymbol{\epsilon})\} = E\{\mathbf{W}_k(\boldsymbol{\epsilon})\} = 0$. If we define $\delta \bar{\mathbf{v}}_l = \bar{\mathbf{v}}_l - \boldsymbol{\nu}_l = \sigma_\eta \mathbf{v}_l^{(1)} + O(\sigma_\eta^2)$, then $E\{\delta \bar{\mathbf{v}}_l\} = \mathbf{0}_p + O(\sigma_\eta^2)$ and $\text{cov}\{\delta \bar{\mathbf{v}}_l, \delta \bar{\mathbf{v}}_m\} = E\{\delta \bar{\mathbf{v}}_l \delta \bar{\mathbf{v}}_m^H\} + O(\sigma_\eta^4)$. But $E\{\delta \bar{\mathbf{v}}_l \delta \bar{\mathbf{v}}_m^H\} = \sigma_\eta^2 E\{\mathbf{v}_l^{(1)} \mathbf{v}_m^{(1)H}\} + O(\sigma_\eta^3)$, so that from (B 11),

$$\begin{aligned} \text{cov}\{\delta \bar{\mathbf{v}}_l, \delta \bar{\mathbf{v}}_m\} &= \frac{\sigma_\eta^2}{K^2} \sum_{r \neq l} \sum_{k=0}^{K-1} \sum_{s \neq m} \sum_{n=0}^{K-1} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_s)^{-1} \times \\ &\quad \boldsymbol{\nu}_r \boldsymbol{\nu}_r^H \left\{ E\{\mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\boldsymbol{\epsilon}) \boldsymbol{\nu}_l \boldsymbol{\nu}_m^H \mathbf{W}_n(\boldsymbol{\epsilon}) \mathbf{W}_n^H(\mathbf{x})\} + \right. \\ &\quad E\{\mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\boldsymbol{\epsilon}) \boldsymbol{\nu}_l \boldsymbol{\nu}_m^H \mathbf{W}_n(\mathbf{x}) \mathbf{W}_n^H(\boldsymbol{\epsilon})\} + \\ &\quad E\{\mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_k^H(\mathbf{x}) \boldsymbol{\nu}_l \boldsymbol{\nu}_m^H \mathbf{W}_n(\boldsymbol{\epsilon}) \mathbf{W}_n^H(\mathbf{x})\} + \\ &\quad \left. E\{\mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_k^H(\mathbf{x}) \boldsymbol{\nu}_l \boldsymbol{\nu}_m^H \mathbf{W}_n(\mathbf{x}) \mathbf{W}_n^H(\boldsymbol{\epsilon})\} \right\} \boldsymbol{\nu}_s \boldsymbol{\nu}_s^H + O(\sigma_\eta^3). \end{aligned} \quad (\text{B } 12)$$

Now, due to (5.4),

$$\begin{aligned} E\{\mathbf{W}_k^H(\boldsymbol{\epsilon}) \boldsymbol{\nu}_l \boldsymbol{\nu}_m^H(\boldsymbol{\epsilon}) \boldsymbol{\nu}_s\} &= \boldsymbol{\nu}_l^T E^* \{\mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_n^T(\boldsymbol{\epsilon})\} \boldsymbol{\nu}_s = 0, \text{ and} \\ E\{\boldsymbol{\nu}_r^H \mathbf{W}_k(\boldsymbol{\epsilon}) \boldsymbol{\nu}_m^H \mathbf{W}_n(\boldsymbol{\epsilon})\} &= \boldsymbol{\nu}_r^H E\{\mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_n^T(\boldsymbol{\epsilon})\} \boldsymbol{\nu}_m^* = 0, \end{aligned}$$

so that the second and third expectations in (B 12) are both zero. By way of contrast, using (5.3), with $*$ denoting complex conjugation,

$$\begin{aligned} E\{\mathbf{W}_k^H(\boldsymbol{\epsilon}) \boldsymbol{\nu}_l \boldsymbol{\nu}_m^H \mathbf{W}_n(\boldsymbol{\epsilon})\} &= \boldsymbol{\nu}_l^T E^* \{\mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_n^H(\boldsymbol{\epsilon})\} \boldsymbol{\nu}_m^* = \boldsymbol{\nu}_l^T \mathbf{I}_p \delta_{k,n} \boldsymbol{\nu}_m^* = \delta_{k,n} \delta_{l,m}, \\ E\{\boldsymbol{\nu}_r^H \mathbf{W}_k(\boldsymbol{\epsilon}) \boldsymbol{\nu}_m^H \mathbf{W}_n(\boldsymbol{\epsilon}) \boldsymbol{\nu}_s\} &= \boldsymbol{\nu}_r^H E\{\mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_n^H(\boldsymbol{\epsilon})\} \boldsymbol{\nu}_s = \boldsymbol{\nu}_r^H \mathbf{I}_p \delta_{k,n} \boldsymbol{\nu}_s = \delta_{k,n} \delta_{r,s}. \end{aligned}$$

Thus,

$$\begin{aligned} \text{cov}\{\delta\bar{\mathbf{v}}_l, \delta\bar{\mathbf{v}}_m\} &= \frac{\sigma_\eta^2}{K^2} \sum_{r \neq l} \sum_{k=0}^{K-1} \sum_{s \neq m} \sum_{n=0}^{K-1} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_s)^{-1} \times \\ &\quad \boldsymbol{\nu}_r \left\{ \delta_{k,n} \delta_{l,m} \boldsymbol{\nu}_r^H \mathbf{W}_k(\mathbf{x}) \mathbf{W}_n^H(\mathbf{x}) \boldsymbol{\nu}_s + \delta_{k,n} \delta_{r,s} \boldsymbol{\nu}_m^H \mathbf{W}_n(\mathbf{x}) \mathbf{W}_k^H(\mathbf{x}) \boldsymbol{\nu}_l \right\} \boldsymbol{\nu}_s^H + O(\sigma_\eta^3). \end{aligned} \quad (\text{B } 13)$$

Then,

$$\begin{aligned} &\frac{\sigma_\eta^2}{K^2} \sum_{r \neq l} \sum_{k=0}^{K-1} \sum_{s \neq m} \sum_{n=0}^{K-1} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_s)^{-1} \boldsymbol{\nu}_r \delta_{k,n} \delta_{l,m} \boldsymbol{\nu}_r^H \mathbf{W}_k(\mathbf{x}) \mathbf{W}_n^H(\mathbf{x}) \boldsymbol{\nu}_s \boldsymbol{\nu}_s^H \\ &= \frac{\sigma_\eta^2}{K} \sum_{r \neq l} \sum_{s \neq m} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_s)^{-1} \boldsymbol{\nu}_r \delta_{l,m} \boldsymbol{\nu}_r^H \mathbf{S}^{(0)} \boldsymbol{\nu}_s \boldsymbol{\nu}_s^H \\ &= \frac{\sigma_\eta^2}{K} \sum_{r \neq l} \sum_{s \neq m} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_s)^{-1} \boldsymbol{\nu}_r \delta_{l,m} \boldsymbol{\nu}_r^H \zeta_s \boldsymbol{\nu}_s \boldsymbol{\nu}_s^H \\ &= \frac{\sigma_\eta^2}{K} \sum_{r \neq l} \sum_{s \neq m} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_s)^{-1} \zeta_s \delta_{l,m} \delta_{r,s} \boldsymbol{\nu}_r \boldsymbol{\nu}_s^H \\ &= \frac{\delta_{l,m} \sigma_\eta^2}{K} \sum_{r \neq l} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_r)^{-1} \zeta_r \boldsymbol{\nu}_r \boldsymbol{\nu}_r^H. \end{aligned}$$

In a similar way we find that,

$$\begin{aligned} &\frac{\sigma_\eta^2}{K^2} \sum_{r \neq l} \sum_{k=0}^{K-1} \sum_{s \neq m} \sum_{n=0}^{K-1} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_s)^{-1} \boldsymbol{\nu}_r \delta_{k,n} \delta_{r,s} \boldsymbol{\nu}_m^H \mathbf{W}_n(\mathbf{x}) \mathbf{W}_k^H(\mathbf{x}) \boldsymbol{\nu}_l \boldsymbol{\nu}_s^H \\ &= \frac{\delta_{l,m} \sigma_\eta^2}{K} \sum_{r \neq l} (\zeta_l - \zeta_r)^{-1} (\zeta_m - \zeta_r)^{-1} \zeta_l \boldsymbol{\nu}_r \boldsymbol{\nu}_r^H, \end{aligned}$$

so that $\text{cov}\{\delta\bar{\mathbf{v}}_l, \delta\bar{\mathbf{v}}_m\} = \sigma_\eta^2 \frac{\delta_{l,m}}{K} \sum_{r \neq l} \frac{(\zeta_r + \zeta_l)}{(\zeta_r - \zeta_l)^2} \boldsymbol{\nu}_r \boldsymbol{\nu}_r^H + O(\sigma_\eta^3)$.

Now $\text{cov}\{\delta\bar{\mathbf{v}}_l, \delta\bar{\mathbf{v}}_m^*\} = E\{\delta\bar{\mathbf{v}}_l \delta\bar{\mathbf{v}}_m^T\} + O(\sigma_\eta^4)$ and $E\{\delta\bar{\mathbf{v}}_l \delta\bar{\mathbf{v}}_m^T\} = \sigma_\eta^2 E\{\mathbf{v}_l^{(1)} \mathbf{v}_l^{(1)T}\} + O(\sigma_\eta^3)$. We find $\text{cov}\{\delta\bar{\mathbf{v}}_l, \delta\bar{\mathbf{v}}_m^*\} = -\sigma_\eta^2 \frac{(1 - \delta_{l,m})}{K} \frac{(\zeta_m + \zeta_l)}{(\zeta_m - \zeta_l)^2} \boldsymbol{\nu}_m \boldsymbol{\nu}_l^T + O(\sigma_\eta^3)$, where we have used similar methods to those above.

Let $\mathbf{v}_l = c\bar{\mathbf{v}}_l$, where $\mathbf{v}_l$ has Euclidean norm unity, and c is a constant. Then $\bar{\mathbf{v}}_l$ has norm of $1/c$. But $\langle \bar{\mathbf{v}}_l, \bar{\mathbf{v}}_l \rangle = 1 + O(\sigma_\eta^2)$, so $1/c = \sqrt{1 + O(\sigma_\eta^2)}$. Hence,

$$\begin{aligned} \text{cov}\{\delta\mathbf{v}_l, \delta\mathbf{v}_m\} &= c^2 \text{cov}\{\delta\bar{\mathbf{v}}_l, \delta\bar{\mathbf{v}}_m\} \\ &= [1 + O(\sigma_\eta^2)] \left[\sigma_\eta^2 \frac{\delta_{l,m}}{K} \sum_{r \neq l} \frac{(\zeta_r + \zeta_l)}{(\zeta_r - \zeta_l)^2} \boldsymbol{\nu}_r \boldsymbol{\nu}_r^H + O(\sigma_\eta^3) \right] \\ &= \sigma_\eta^2 \frac{\delta_{l,m}}{K} \sum_{r \neq l} \frac{(\zeta_r + \zeta_l)}{(\zeta_r - \zeta_l)^2} \boldsymbol{\nu}_r \boldsymbol{\nu}_r^H + O(\sigma_\eta^3). \end{aligned} \quad (\text{B } 14)$$

Similarly we obtain

$$\text{cov}\{\delta\mathbf{v}_l, \delta\mathbf{v}_m^*\} = -\sigma_\eta^2 \frac{(1 - \delta_{l,m})}{K} \frac{(\zeta_m + \zeta_l)}{(\zeta_m - \zeta_l)^2} \boldsymbol{\nu}_m \boldsymbol{\nu}_l^T + O(\sigma_\eta^3). \quad (\text{B } 15)$$

We note that (B 15) gives a zero covariance between $\delta \mathbf{v}_l$ and $\delta \mathbf{v}_m^*$ when $l = m$. The first terms on the right of (B 14) and (B 15) agree with Park and Chave (1984, eqns. (2.18) and (2.19)), although their results arose in a quite different context.

(b) *Zero Eigenvalues and Estimation of Noise Variance*

Here we derive an estimate of σ_η^2 using the idea that some of the small but non-zero eigenvalues of the noisy matrix $\hat{\mathbf{S}}(\mathbf{X})$ correspond to identically zero eigenvalues of the noise-free matrix $\mathbf{S}^{(0)}$. Our approach is based on that of Backus *et al* (1981, pp. 422–6). We again aim for a clear but precise formulation. (Some steps differ completely in methodology.) The first term on the right of (B 26) agrees with the result in Backus *et al* (1981, p. 427).

Although we expect all the eigenvalues $\{\lambda_l = \xi_l^2, l = 1, \dots, p\}$ of $\hat{\mathbf{S}}(\mathbf{X}) = \mathbf{M}^H \mathbf{M}$ to be non-zero because of the noise, the eigenvalues $\{\zeta_l, l = 1, \dots, p\}$ of the noise-free matrix $\mathbf{S}^{(0)}$ need not all be non-zero. With the eigenvalues ranked in descending order of magnitude, suppose $\zeta_{p-q+1}, \dots, \zeta_p = 0$, i.e., the last q eigenvalues are identically zero. Let $l \in \{p - q + 1, \dots, p\}$. Then from (B 8)

$$\mathbf{S}^{(0)} \mathbf{v}_l^{(1)} = (\lambda_l^{(1)} \mathbf{I}_p - \mathbf{S}^{(1)}) \mathbf{v}_l. \quad (\text{B } 16)$$

Now, $\mathbf{S}^{(0)} = \sum_{k=1}^{p-q} \zeta_k \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H$ and we can define a projection operator which projects vectors to the linear subspace consisting of all eigenvectors of $\mathbf{S}^{(0)}$ with zero eigenvalue, $\mathbf{P}^{(0)} = \sum_{j=p-q+1}^p \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H$, for which

$$\mathbf{S}^{(0)} \mathbf{P}^{(0)} = \mathbf{P}^{(0)} \mathbf{S}^{(0)} = \sum_{k=1}^{p-q} \zeta_k \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H \sum_{j=p-q+1}^p \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H = \sum_{k=1}^{p-q} \sum_{j=p-q+1}^p \zeta_k \boldsymbol{\nu}_k \boldsymbol{\nu}_k^H \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H = \mathbf{0}_{p \times p},$$

due to the orthonormality of the $\{\boldsymbol{\nu}_k\}$. Multiplying (B 16) by $\mathbf{P}^{(0)}$ we then see that

$$\mathbf{P}^{(0)} (\lambda_l^{(1)} \mathbf{I}_p - \mathbf{S}^{(1)}) \mathbf{v}_l = \mathbf{0}_p. \quad (\text{B } 17)$$

Also

$$\mathbf{P}^{(0)} \mathbf{v}_l = \sum_{j=p-q+1}^p \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H \mathbf{v}_l = \sum_{j=p-q+1}^p \boldsymbol{\nu}_j \delta_{j,l} = \mathbf{v}_l. \quad (\text{B } 18)$$

Further we have that

$$\begin{aligned} \mathbf{P}^{(0)} \mathbf{S}^{(1)} \mathbf{v}_l &= \frac{1}{K} \sum_{j=p-q+1}^p \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H \sum_{k=0}^{K-1} [\mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\boldsymbol{\epsilon}) + \mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_k^H(\mathbf{x})] \mathbf{v}_l \\ &= \frac{1}{K} \sum_{j=p-q+1}^p \sum_{k=0}^{K-1} \boldsymbol{\nu}_j [\boldsymbol{\nu}_j^H \mathbf{W}_k(\mathbf{x})] \mathbf{W}_k^H(\boldsymbol{\epsilon}) \mathbf{v}_l \\ &\quad + \frac{1}{K} \sum_{j=p-q+1}^p \sum_{k=0}^{K-1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H \mathbf{W}_k(\boldsymbol{\epsilon}) [\mathbf{W}_k^H(\mathbf{x}) \mathbf{v}_l] = \mathbf{0}_p, \end{aligned} \quad (\text{B } 19)$$

since the terms in square brackets are zero. To see this note that

$$\begin{aligned}
0 &= \text{tr}\{\mathbf{0}_{p \times p}\} = \text{tr}\{\mathbf{P}^{(0)}\mathbf{S}^{(0)}\} = \text{tr}\left\{\sum_{j=p-q+1}^p \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H \frac{1}{K} \sum_{k=0}^{K-1} \mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\mathbf{x})\right\} \\
&= \frac{1}{K} \sum_{j=p-q+1}^p \sum_{k=0}^{K-1} \text{tr}\{\boldsymbol{\nu}_j \langle \boldsymbol{\nu}_j, \mathbf{W}_k(\mathbf{x}) \rangle \mathbf{W}_k^H(\mathbf{x})\} \\
&= \frac{1}{K} \sum_{j=p-q+1}^p \sum_{k=0}^{K-1} \langle \boldsymbol{\nu}_j, \mathbf{W}_k(\mathbf{x}) \rangle \text{tr}\{\langle \mathbf{W}_k(\mathbf{x}), \boldsymbol{\nu}_j \rangle\} \\
&= \frac{1}{K} \sum_{j=p-q+1}^p \sum_{k=0}^{K-1} |\langle \mathbf{W}_k(\mathbf{x}), \boldsymbol{\nu}_j \rangle|^2 \\
&\Rightarrow \mathbf{W}_k^H(\mathbf{x}) \boldsymbol{\nu}_j = 0, \quad k \in \{0, \dots, K-1\}, \quad j \in \{p-q+1, \dots, p\}. \tag{B 20}
\end{aligned}$$

Then from (B 17), (B 18) and (B 19) we get

$$\lambda_l^{(1)} \mathbf{P}^{(0)} \boldsymbol{\nu}_l = \lambda_l^{(1)} \boldsymbol{\nu}_l = \mathbf{P}^{(0)} \mathbf{S}^{(1)} \boldsymbol{\nu}_l = \mathbf{0}_p, \tag{B 21}$$

so that $\lambda_l^{(1)} \boldsymbol{\nu}_l^H \boldsymbol{\nu}_l = \lambda_l^{(1)} = 0$ for all $l \in \{p-q+1, \dots, p\}$. Then, for these values of l , equating terms of $O(\sigma_\eta^2)$ in (B 7) gives,

$$\begin{aligned}
\mathbf{S}^{(0)} \mathbf{v}_l^{(2)} + \mathbf{S}^{(1)} \mathbf{v}_l^{(1)} + \mathbf{S}^{(2)} \boldsymbol{\nu}_l &= \lambda_l^{(2)} \boldsymbol{\nu}_l \\
\Rightarrow \mathbf{P}^{(0)} \mathbf{S}^{(0)} \mathbf{v}_l^{(2)} + \mathbf{P}^{(0)} \mathbf{S}^{(1)} \mathbf{v}_l^{(1)} + \mathbf{P}^{(0)} \mathbf{S}^{(2)} \boldsymbol{\nu}_l &= \mathbf{P}^{(0)} \lambda_l^{(2)} \boldsymbol{\nu}_l = \lambda_l^{(2)} \boldsymbol{\nu}_l \\
\text{i.e., } \mathbf{P}^{(0)} \mathbf{S}^{(1)} \mathbf{v}_l^{(1)} + \mathbf{P}^{(0)} \mathbf{S}^{(2)} \boldsymbol{\nu}_l &= \lambda_l^{(2)} \boldsymbol{\nu}_l, \tag{B 22}
\end{aligned}$$

since $\mathbf{P}^{(0)} \mathbf{S}^{(0)} = \mathbf{0}_{p \times p}$ and $\mathbf{P}^{(0)} \boldsymbol{\nu}_l = \boldsymbol{\nu}_l$. Using analogous arguments to those used in deriving (B 10), for the current case of degenerate zero-valued eigenvalues gives $\mathbf{v}_l^{(1)} = \sum_{j=p-q+1}^p \boldsymbol{\nu}_j \langle \boldsymbol{\nu}_j, \mathbf{v}_l^{(1)} \rangle - (\mathbf{S}^{(0)})^+ \mathbf{S}^{(1)} \boldsymbol{\nu}_l$, where $(\mathbf{S}^{(0)})^+ = \sum_{m=1}^{p-q} \zeta_m^{-1} \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H$, so that, from (B 21),

$$\begin{aligned}
\mathbf{P}^{(0)} \mathbf{S}^{(1)} \mathbf{v}_l^{(1)} &= \mathbf{P}^{(0)} \mathbf{S}^{(1)} \left[\sum_{j=p-q+1}^p \boldsymbol{\nu}_j \langle \boldsymbol{\nu}_j, \mathbf{v}_l^{(1)} \rangle - (\mathbf{S}^{(0)})^+ \mathbf{S}^{(1)} \boldsymbol{\nu}_l \right] \\
&= \mathbf{0}_p - \mathbf{P}^{(0)} \mathbf{S}^{(1)} (\mathbf{S}^{(0)})^+ \mathbf{S}^{(1)} \boldsymbol{\nu}_l,
\end{aligned}$$

so (B 22) becomes the eigenequation $\mathbf{P}^{(0)} [\mathbf{S}^{(2)} - \mathbf{S}^{(1)} (\mathbf{S}^{(0)})^+ \mathbf{S}^{(1)}] \boldsymbol{\nu}_l = \lambda_l^{(2)} \boldsymbol{\nu}_l$, and since $\mathbf{P}^{(0)} \boldsymbol{\nu}_l = \boldsymbol{\nu}_l$ we get $\mathbf{P}^{(0)} [\mathbf{S}^{(2)} - \mathbf{S}^{(1)} (\mathbf{S}^{(0)})^+ \mathbf{S}^{(1)}] \mathbf{P}^{(0)} \boldsymbol{\nu}_l = \lambda_l^{(2)} \boldsymbol{\nu}_l$, which, with $\tilde{\mathbf{S}}^{(2)} \equiv \mathbf{P}^{(0)} [\mathbf{S}^{(2)} - \mathbf{S}^{(1)} (\mathbf{S}^{(0)})^+ \mathbf{S}^{(1)}] \mathbf{P}^{(0)}$ gives

$$\tilde{\mathbf{S}}^{(2)} \boldsymbol{\nu}_l = \lambda_l^{(2)} \boldsymbol{\nu}_l, \quad l \in \{p-q+1, \dots, p\}. \tag{B 23}$$

But $\tilde{\mathbf{S}}^{(2)} - \mathbf{P}^{(0)}\mathbf{S}^{(2)}\mathbf{P}^{(0)} = -\mathbf{P}^{(0)}\mathbf{S}^{(1)}(\mathbf{S}^{(0)}) + \mathbf{S}^{(1)}\mathbf{P}^{(0)}$ and this last expression equals

$$\begin{aligned} & -\frac{1}{K^2} \sum_{j,n=p-q+1}^p \sum_{i,k=0}^{K-1} \sum_{m=1}^{p-q} \boldsymbol{\nu}_n \boldsymbol{\nu}_n^H [\mathbf{W}_i(\mathbf{x}) \mathbf{W}_i^H(\boldsymbol{\epsilon}) + \mathbf{W}_i(\boldsymbol{\epsilon}) \mathbf{W}_i^H(\mathbf{x})] \times \\ & \quad \zeta_m^{-1} \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H [\mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\boldsymbol{\epsilon}) + \mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_k^H(\mathbf{x})] \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H \\ & = -\frac{1}{K^2} \sum_{j,n,i,k,m} \boldsymbol{\nu}_n \boldsymbol{\nu}_n^H \mathbf{W}_i(\boldsymbol{\epsilon}) \mathbf{W}_i^H(\mathbf{x}) \zeta_m^{-1} \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H \mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\boldsymbol{\epsilon}) \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H, \\ & = -\frac{1}{\zeta_m K^2} \sum_{j,n,i,k,m} \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \langle \boldsymbol{\nu}_n, \mathbf{W}_i(\boldsymbol{\epsilon}) \rangle \langle \mathbf{W}_i(\mathbf{x}), \boldsymbol{\nu}_m \rangle \langle \boldsymbol{\nu}_m, \mathbf{W}_k(\mathbf{x}) \rangle \langle \mathbf{W}_k(\boldsymbol{\epsilon}), \boldsymbol{\nu}_j \rangle, \end{aligned}$$

other terms being zero due to the result (B 20). Now

$$\begin{aligned} \mathbf{P}^{(0)}\mathbf{S}^{(2)}\mathbf{P}^{(0)} & = \frac{1}{K} \sum_{j,n=p-q+1}^p \sum_{k=0}^{K-1} \boldsymbol{\nu}_n \boldsymbol{\nu}_n^H \mathbf{W}_k(\boldsymbol{\epsilon}) \mathbf{W}_k^H(\boldsymbol{\epsilon}) \boldsymbol{\nu}_j \boldsymbol{\nu}_j^H \\ & = \frac{1}{K} \sum_{j,n,k} \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \langle \boldsymbol{\nu}_n, \mathbf{W}_k(\boldsymbol{\epsilon}) \rangle \langle \mathbf{W}_k(\boldsymbol{\epsilon}), \boldsymbol{\nu}_j \rangle \\ & = \frac{1}{K} \sum_{j,n=p-q+1}^p \sum_{i,k=0}^{K-1} \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \langle \boldsymbol{\nu}_n, \mathbf{W}_i(\boldsymbol{\epsilon}) \rangle \langle \mathbf{W}_k(\boldsymbol{\epsilon}), \boldsymbol{\nu}_j \rangle \delta_{i,k}. \end{aligned}$$

We can thus express $\tilde{\mathbf{S}}^{(2)}$ as

$$\begin{aligned} \tilde{\mathbf{S}}^{(2)} & = \frac{1}{K} \sum_{j,n=p-q+1}^p \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \sum_{i,k=0}^{K-1} \langle \boldsymbol{\nu}_n, \mathbf{W}_i(\boldsymbol{\epsilon}) \rangle \langle \mathbf{W}_k(\boldsymbol{\epsilon}), \boldsymbol{\nu}_j \rangle \times \\ & \quad \left\{ \delta_{i,k} - \frac{1}{\zeta_m K} \sum_{m=1}^{p-q} \langle \mathbf{W}_i(\mathbf{x}), \boldsymbol{\nu}_m \rangle \langle \boldsymbol{\nu}_m, \mathbf{W}_k(\mathbf{x}) \rangle \right\}. \end{aligned}$$

(agreeing with Backus, Park and Garbasz 1981, eqn. (27b)).

If we define the $(p-q) \times K$ matrix $\mathbf{G}$ by $(\mathbf{G})_{m,k} = [1/\sqrt{(\zeta_m K)}] \langle \boldsymbol{\nu}_m, \mathbf{W}_k(\mathbf{x}) \rangle = G_{mk}$, say, for $m = 1, \dots, p-q$ and $k = 0, \dots, K-1$, then it is easy to show it is unitary:

$$\begin{aligned} (\mathbf{G}\mathbf{G}^H)_{ij} & = \sum_{k=0}^{K-1} G_{ik} G_{jk}^H = \frac{1}{K\sqrt{(\zeta_i \zeta_j)}} \sum_{k=0}^{K-1} \langle \boldsymbol{\nu}_i, \mathbf{W}_k(\mathbf{x}) \rangle \langle \boldsymbol{\nu}_j, \mathbf{W}_k(\mathbf{x}) \rangle^H \\ & = \frac{1}{K\sqrt{(\zeta_i \zeta_j)}} \boldsymbol{\nu}_i^H \sum_{k=0}^{K-1} \mathbf{W}_k(\mathbf{x}) \mathbf{W}_k^H(\mathbf{x}) \boldsymbol{\nu}_j \\ & = \frac{1}{\sqrt{(\zeta_i \zeta_j)}} \boldsymbol{\nu}_i^H \mathbf{S}^{(0)} \boldsymbol{\nu}_j = \sqrt{(\zeta_j / \zeta_i)} \boldsymbol{\nu}_i^H \boldsymbol{\nu}_j = \delta_{i,j}. \end{aligned}$$

We can do a singular value decomposition of $\mathbf{G}$ as $\mathbf{G} = \mathbf{A}\boldsymbol{\Theta}\mathbf{C}^H$, where $\mathbf{A}$ is a $(p-q) \times (p-q)$ unitary matrix, $\mathbf{C}$ is a $K \times K$ unitary matrix, and $\boldsymbol{\Theta}$ is a diagonal

$(p - q) \times K$ matrix of singular values,

$$\boldsymbol{\Theta} = \begin{bmatrix} \theta_1 & & & \\ & \ddots & & \\ & & \theta_{p-q} & \\ & & & \mathbf{0}_{(p-q) \times (K-p+q)} \end{bmatrix}.$$

Then since both $\mathbf{G}$ and $\mathbf{C}$ are unitary, $\mathbf{G}\mathbf{G}^H = \mathbf{I}_{p-q} = \mathbf{A}\boldsymbol{\Theta}\mathbf{C}^H\mathbf{C}\boldsymbol{\Theta}^H\mathbf{A}^H = \mathbf{A}\boldsymbol{\Theta}\boldsymbol{\Theta}^H\mathbf{A}^H \Rightarrow \boldsymbol{\Theta}\boldsymbol{\Theta}^H = \mathbf{I}_{p-q}$, so that the singular values θ_i all have absolute value unity; this must also hold because $\theta_1, \dots, \theta_{p-q}$ are the square roots of the eigenvalues of $\mathbf{G}\mathbf{G}^H$ which is unitary, and every eigenvalue of a unitary matrix has absolute value unity (Strang 1988, p. 298). Then

$$\begin{aligned} \mathbf{G}^H\mathbf{G} &= \mathbf{C}\boldsymbol{\Theta}^H\mathbf{A}^H\mathbf{A}\boldsymbol{\Theta}\mathbf{C}^H = \mathbf{C}\boldsymbol{\Theta}^H\boldsymbol{\Theta}\mathbf{C}^H \\ &= \mathbf{C} \begin{bmatrix} \mathbf{I}_{p-q} & \mathbf{0}_{(p-q) \times (K-p+q)} \\ \mathbf{0}_{(K-p+q) \times (p-q)} & \mathbf{0}_{(K-p+q) \times (K-p+q)} \end{bmatrix} \mathbf{C}^H. \end{aligned}$$

$$\text{where } (\mathbf{G}^H\mathbf{G})_{ij} = \sum_{m=1}^{p-q} G_{mi}^H G_{mj} = \frac{1}{K\zeta_m} \sum_{m=1}^{p-q} \langle \mathbf{W}_i(\mathbf{x}), \boldsymbol{\nu}_m \rangle \langle \boldsymbol{\nu}_m, \mathbf{W}_j(\mathbf{x}) \rangle.$$

So,

$$\begin{aligned} \tilde{\mathbf{S}}^{(2)} &= \frac{1}{K} \sum_{j,n=p-q+1}^p \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \sum_{i,k=0}^{K-1} \langle \boldsymbol{\nu}_n, \mathbf{W}_i(\boldsymbol{\epsilon}) \rangle \langle \mathbf{W}_k(\boldsymbol{\epsilon}), \boldsymbol{\nu}_j \rangle \times \\ &\quad \left\{ \delta_{i,k} - \left(\mathbf{C} \begin{bmatrix} \mathbf{I}_{p-q} & \mathbf{0}_{(p-q) \times (K-p+q)} \\ \mathbf{0}_{(K-p+q) \times (p-q)} & \mathbf{0}_{(K-p+q) \times (K-p+q)} \end{bmatrix} \mathbf{C}^H \right)_{ik} \right\} \\ &= \frac{1}{K} \sum_{j,n=p-q+1}^p \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \sum_{i,k=0}^{K-1} \langle \boldsymbol{\nu}_n, \mathbf{W}_i(\boldsymbol{\epsilon}) \rangle \langle \mathbf{W}_k(\boldsymbol{\epsilon}), \boldsymbol{\nu}_j \rangle \times \\ &\quad \left\{ \left(\mathbf{C} \begin{bmatrix} \mathbf{0}_{(p-q) \times (p-q)} & \mathbf{0}_{(p-q) \times (K-p+q)} \\ \mathbf{0}_{(K-p+q) \times (p-q)} & \mathbf{I}_{(K-p+q)} \end{bmatrix} \mathbf{C}^H \right)_{ik} \right\} \\ &= \frac{1}{K} \sum_{j,n=p-q+1}^p \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \sum_{i,k=0}^{K-1} \langle \boldsymbol{\nu}_n, \mathbf{W}_i(\boldsymbol{\epsilon}) \rangle \langle \mathbf{W}_k(\boldsymbol{\epsilon}), \boldsymbol{\nu}_j \rangle \sum_{r=p-q}^{K-1} c_{ir} c_{kr}^H, \end{aligned}$$

since the term in the curly bracket is $\sum_{r,s=p-q}^{K-1} c_{ir} \delta_{r,s} c_{ks}^H = \sum_{r=p-q}^{K-1} c_{ir} c_{kr}^H$. Now define

$$F_{ij} = \langle \boldsymbol{\nu}_i, \mathbf{W}_j(\boldsymbol{\epsilon}) \rangle ; y_{ij} = \sum_{k=0}^{K-1} F_{ik} c_{kj} \text{ and } \mathbf{q}_i = \sum_{j=p-q+1}^p \boldsymbol{\nu}_j y_{ji}.$$

Then

$$\begin{aligned}
\tilde{\mathbf{S}}^{(2)} &= \frac{1}{K} \sum_{j,n=p-q+1}^p \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \sum_{i,k=0}^{K-1} F_{ni} F_{jk}^H \sum_{r=p-q}^{K-1} c_{ir} c_{kr}^H \\
&= \frac{1}{K} \sum_{r=p-q}^{K-1} \sum_{j,n=p-q+1}^p \boldsymbol{\nu}_n \boldsymbol{\nu}_j^H \sum_{i=0}^{K-1} F_{ni} c_{ir} \sum_{k=0}^{K-1} F_{jk}^H c_{kr}^H \\
&= \frac{1}{K} \sum_{r=p-q}^{K-1} \sum_{n=p-q+1}^p \boldsymbol{\nu}_n y_{nr} \sum_{j=p-q+1}^p \boldsymbol{\nu}_j^H y_{jr}^H = \frac{1}{K} \sum_{r=p-q}^{K-1} \mathbf{q}_r \mathbf{q}_r^H.
\end{aligned}$$

The statistical properties of $\tilde{\mathbf{S}}^{(2)}$ thus depend on those of the $\{\mathbf{q}_r\}$, which can be gleaned after the following intermediate steps.

$$\begin{aligned}
E\{F_{ij}\} &= \boldsymbol{\nu}_i^H E\{\mathbf{W}_j(\boldsymbol{\epsilon})\} = 0; \quad E\{F_{ij} F_{mn}\} = \boldsymbol{\nu}_i^H E\{\mathbf{W}_j(\boldsymbol{\epsilon}) \mathbf{W}_n^T(\boldsymbol{\epsilon})\} \boldsymbol{\nu}_m^* = 0, \\
E\{F_{ij} F_{mn}^H\} &= \boldsymbol{\nu}_i^H E\{\mathbf{W}_j(\boldsymbol{\epsilon}) \mathbf{W}_n^H(\boldsymbol{\epsilon})\} \boldsymbol{\nu}_m = \boldsymbol{\nu}_i^H \mathbf{I}_p \delta_{j,n} \boldsymbol{\nu}_m = \delta_{j,n} \delta_{i,m},
\end{aligned}$$

and

$$E\{y_{ij}\} = \sum_{k=0}^{K-1} E\{F_{ik}\} c_{kj} = 0; \quad E\{y_{ij} y_{mn}\} = \sum_{k,l=0}^{K-1} E\{F_{ik} F_{ml}\} c_{kj} c_{ln} = 0, \quad (\text{B } 24)$$

$$E\{y_{ij} y_{mn}^H\} = \sum_{k,l=0}^{K-1} E\{F_{ik} F_{ml}^H\} c_{kj} c_{ln}^H = \delta_{i,m} \sum_{k,l=0}^{K-1} \delta_{k,l} c_{kj} c_{ln}^H = \delta_{i,m} \delta_{j,n}, \quad (\text{B } 25)$$

since $\mathbf{C}$ is unitary. It then follows that $E\{\mathbf{q}_r\} = \sum_{j=p-q+1}^p \boldsymbol{\nu}_j E\{y_{jr}\} = \mathbf{0}_p$, and

$$E\{\mathbf{q}_r \mathbf{q}_s^H\} = \sum_{j,l=p-q+1}^p \boldsymbol{\nu}_j \boldsymbol{\nu}_l^H E\{y_{jr} y_{lr}^H\} = \sum_{j,l=p-q+1}^p \boldsymbol{\nu}_j \boldsymbol{\nu}_l^H \delta_{j,l} \delta_{r,s} = \delta_{r,s} \sum_{l=p-q+1}^p \boldsymbol{\nu}_l \boldsymbol{\nu}_l^H = \mathbf{P}^{(0)} \delta_{r,s}.$$

Let $\mathcal{V}$ be the $p \times p$ unitary matrix with columns $\boldsymbol{\nu}_l$. Then $\tilde{\mathbf{S}}^{(2)}$ and $\mathcal{V}^H \tilde{\mathbf{S}}^{(2)} \mathcal{V}$ are similar and so have the same eigenvalues. But

$$\mathcal{V}^H \tilde{\mathbf{S}}^{(2)} \mathcal{V} = \frac{1}{K} \sum_{r=p-q}^{K-1} \mathcal{V}^H \mathbf{q}_r \mathbf{q}_r^H \mathcal{V} = \frac{1}{K} \sum_{r=p-q}^{K-1} \tilde{\mathbf{q}}_r \tilde{\mathbf{q}}_r^H,$$

where

$$\tilde{\mathbf{q}}_r = \mathcal{V}^H \mathbf{q}_r = \begin{bmatrix} \boldsymbol{\nu}_1^H \\ \vdots \\ \boldsymbol{\nu}_p^H \end{bmatrix} \mathbf{q}_r = \sum_{j=p-q+1}^p \begin{bmatrix} \boldsymbol{\nu}_1^H \\ \vdots \\ \boldsymbol{\nu}_p^H \end{bmatrix} \boldsymbol{\nu}_j y_{jr} = \sum_{j=p-q+1}^p \mathbf{e}_j y_{jr},$$

and $\mathbf{e}_j$ is the $p \times 1$ vector with unity in the j th position and zeros elsewhere. Then

$$\mathcal{V}^H \tilde{\mathbf{S}}^{(2)} \mathcal{V} = \frac{1}{K} \sum_{r=p-q}^{K-1} \sum_{j=p-q+1}^p \mathbf{e}_j y_{jr} \sum_{l=p-q+1}^p \mathbf{e}_l^H y_{lr}^H = \sum_{j,l=p-q+1}^p \mathbf{e}_j \mathbf{e}_l^H \frac{1}{K} \sum_{r=p-q}^{K-1} y_{jr} y_{lr}^H,$$

so that

$$(\mathcal{V}^H \tilde{\mathbf{S}}^{(2)} \mathcal{V})_{mn} = \begin{cases} \frac{1}{K} \sum_{r=p-q}^{K-1} y_{mr} y_{nr}^H, & \text{for } p-q+1 \leq m, n \leq p; \\ 0, & \text{otherwise.} \end{cases}$$

But since from (B 23), $\tilde{\mathbf{S}}^{(2)} \boldsymbol{\nu}_l = \lambda_l^{(2)} \boldsymbol{\nu}_l$, $l \in \{p-q+1, \dots, p\}$, we also know that

$$(\mathcal{V}^H \tilde{\mathbf{S}}^{(2)} \mathcal{V})_{mn} = \begin{cases} \lambda_n^{(2)} \boldsymbol{\nu}_m^H \boldsymbol{\nu}_n, & \text{for } p-q+1 \leq m, n \leq p; \\ 0, & \text{otherwise.} \end{cases}$$

So

$$\text{tr}\{\tilde{\mathbf{S}}^{(2)}\} = \text{tr}\{\mathcal{V}^H \tilde{\mathbf{S}}^{(2)} \mathcal{V}\} = \frac{1}{K} \sum_{n=p-q+1}^p \sum_{r=p-q}^{K-1} |y_{nr}|^2 = \sum_{n=p-q+1}^p \lambda_n^{(2)}.$$

Now (B 24) and (B 25) imply

$$\begin{aligned} E\{\Re\{y_{ij}\}\Re\{y_{mn}\}\} &= E\{\Im\{y_{ij}\}\Im\{y_{mn}\}\} = \frac{1}{2}\Re\{\delta_{i,m}\delta_{j,n}\} = \frac{1}{2}\delta_{i,m}\delta_{j,n} \text{ and} \\ E\{\Re\{y_{ij}\}\Im\{y_{mn}\}\} &= -E\{\Re\{y_{mn}\}\Im\{y_{ij}\}\} = -\frac{1}{2}\Im\{\delta_{i,m}\delta_{j,n}\} = 0, \text{ for all } i, j, m, n. \end{aligned}$$

Thus

$$\frac{1}{K} \sum_{n=p-q+1}^p \sum_{r=p-q}^{K-1} |y_{nr}|^2 = \frac{1}{K} \sum_{n=p-q+1}^p \sum_{r=p-q}^{K-1} [\Re^2\{y_{nr}\} + \Im^2\{y_{nr}\}],$$

where $E\{y_{nr}\} = 0$, $E\{|y_{nr}|^2\} = \text{var}\{y_{nr}\} = 1$, and $\text{var}\{\Re\{y_{nr}\}\} = \text{var}\{\Im\{y_{nr}\}\} = 1/2$ and the $\Re\{y_{nr}\}$ and $\Im\{y_{nr}\}$ terms in the double sum are all independent (they are uncorrelated and multivariate Gaussian). It then follows that $\Lambda \equiv \sum_{n=p-q+1}^p \lambda_n^{(2)}$ will be distributed as $\frac{1}{2K} \chi_{2q(K-p+q)}^2$, where χ_m^2 is the chi-squared random variable with m degrees of freedom. The random variable $D = \sigma_\eta^2 \Lambda$ has a probability density

$$f_D(d|\sigma_\eta^2) \propto (1/d) (2Kd/\sigma_\eta^2)^{q(K-p+q)} \exp(-Kd/\sigma_\eta^2), \quad d \geq 0,$$

which is maximised at $\hat{\sigma}_\eta^2 = Kd/[q(K-p+q)]$. But $\lambda_l = \zeta_l + \sigma_\eta \lambda_l^{(1)} + \sigma_\eta^2 \lambda_l^{(2)} + O(\sigma_\eta^3) = \sigma_\eta^2 \lambda_l^{(2)} + O(\sigma_\eta^3)$, for $l \in \{p-q+1, \dots, p\}$, since ζ_l and $\lambda_l^{(1)}$ are both zero. Thus $d = \sigma_\eta^2 \sum_{n=p-q+1}^p \lambda_n^{(2)} = \sum_{n=p-q+1}^p \lambda_n + O(\sigma_\eta^3)$, giving

$$\hat{\sigma}_\eta^2 = \frac{K}{q(K-p+q)} \sum_{n=p-q+1}^p \lambda_n + O(\sigma_\eta^3). \quad (\text{B } 26)$$

(c) Covariance Estimation

From (B 14) and (B 26) we have, to $O(\sigma_\eta^3)$, that

$$\text{cov}\{\boldsymbol{\delta}\mathbf{v}_l, \boldsymbol{\delta}\mathbf{v}_m\} = \sigma_\eta^2 \frac{\delta_{l,m}}{K} \sum_{r \neq l} \frac{(\zeta_r + \zeta_l)}{(\zeta_r - \zeta_l)^2} \boldsymbol{\nu}_r \boldsymbol{\nu}_r^H \text{ and } \hat{\sigma}_\eta^2 = \frac{K}{q(K-p+q)} \sum_{n=p-q+1}^p \lambda_n.$$

An estimate of $\text{cov}\{\delta\mathbf{v}_l, \delta\mathbf{v}_m\}$ thus follows by replacing the true variance σ_η^2 by the estimate $\hat{\sigma}_\eta^2$ and the eigenvalues ζ_l and eigenvectors $\mathbf{v}_l$ of the noise-free matrix by the eigenvalues λ_l and eigenvectors $\mathbf{v}_l$ of the noisy matrix. Thus we get,

$$\widehat{\text{cov}}\{\delta\mathbf{v}_l, \delta\mathbf{v}_m\} = \frac{\delta_{l,m}}{q(K-p+q)} \sum_{n=p-q+1}^p \lambda_n \sum_{r \neq l} \frac{(\lambda_r + \lambda_l)}{(\lambda_r - \lambda_l)^2} \mathbf{v}_r \mathbf{v}_r^H.$$

For example, let $l = m = 1$ and $p = 3$ and $q = 2$; then the first eigenvector only represents ‘signal’ and the latter two represent ‘noise’ and $\lambda_1 \gg \lambda_2, \lambda_3$ with $\lambda_2, \lambda_3 \approx 0$, then since $(\lambda_2 + \lambda_1)/(\lambda_2 - \lambda_1)^2 \approx 1/\lambda_1$ and $(\lambda_3 + \lambda_1)/(\lambda_3 - \lambda_1)^2 \approx 1/\lambda_1$,

$$\widehat{\text{cov}}\{\delta\mathbf{v}_1, \delta\mathbf{v}_1\} \approx \frac{\lambda_2 + \lambda_3}{2(K-1)\lambda_1} [\mathbf{v}_2 \mathbf{v}_2^H + \mathbf{v}_3 \mathbf{v}_3^H]. \quad (\text{B } 27)$$

References

- Backus, G., Park, J. and Garbasz, D. 1981 On the relative importance of the driving forces of plate motion. *Geophys. J. R. Astr. Soc.*, **67**, 415–435.
- Bayram, M. & Baraniuk, R. 2000 Multiple window time-varying spectrum estimation, in *Nonlinear and Nonstationary Signal Processing*, W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young, Eds., Cambridge University Press, Cambridge, 292–316.
- Bear, L.K. & Pavlis, G.L. 1997 Estimation of slowness vectors and their uncertainties using multi-wavelet seismic array processing. *Bull. Seism. Soc. Am.* **87**, 755–769.
- Brillinger, D.R. 1981 *Time series, data analysis and theory. Expanded edition*. New York: McGraw-Hill Inc.
- Çakrak, F. & Loughlin, P.J. 2001 Multiwindow time-varying spectrum with instantaneous bandwidth and frequency constraints. *IEEE Trans. Signal Processing* **49**, 1656–1666.
- Daubechies, I. 1988 Time-frequency localization operators: a geometric phase space approach. *IEEE Trans. Information Theory* **34**, 605–612.
- Daubechies, I. & Paul, T. 1988 Time-frequency localisation operators — a geometric phase space approach: II. The use of dilations. *Inverse Problems* **4**, 661–680.
- Kato, T. 1966 *Perturbation theory for linear operators*. New York: Springer-Verlag.
- Lilly, J.M. & Park, J. 1995 Multiwavelet spectral and polarization analyses of seismic records. *Geophys. J. Int.*, **122**, 1001–1021.
- Lucek, E.A. & Balogh, A. 1998 Analysis of the waves associated with the unusual interplanetary shock observed on day 109, 1992. *J. Geophys. Res.*, **103**, 29633–29642.
- Mallat, S. 1998 *A Wavelet Tour of Signal Processing*. San Diego: Academic Press.
- Olhede, S. & Walden, A.T. 2002a Generalized Morse wavelets. *IEEE Trans. Signal Processing*, to appear.
- Olhede, S. & Walden, A.T. 2002b Polarization phase relationships via multiple Morse wavelets — Part II: Data analysis. *Proc. Roy. Soc. Lond. Ser. A*, submitted.
- Park, J. & Chave, A.D. 1984 On the estimation of magnetotelluric response functions using the singular value decomposition. *Geophys. J. R. Astr. Soc.*, **77**, 683–709.
- Park, J., Vernon III, F.L., & Lindberg, C.R. 1987 Frequency-dependent polarization of high-frequency seismograms. *J. Geophys. Res.* **92**, 12 664–12 674.
- Poletti, M.A. 1997 The homomorphic analytic signal. *IEEE Trans. Signal Processing* **45**, 1943–1953.
- Strang, G. 1988 *Linear algebra and its applications. 3rd. edition*. Fort Worth: Harcourt Brace Jovanovich College Publishers.

- Vakman, D. 1967 Optimum signals which maximize partial volume under an ambiguity surface. *Radio Eng. Electron. Phys.* **Aug.**, 1260–1268.
- Wilkinson, J.H. 1965 *The algebraic eigenvalue problem*. Oxford: Oxford University Press.
- Zanandrea, A., Neto, C.M., Rosa, R.R., & Ramos, F.F. 2000 Analysis of geomagnetic pulsations using multiwavelets spectral and polarization method. *Physica A* **283**, 175–180.

1:7:2002