Tapering Promotes Propriety for Fourier Transforms of Real-Valued Time Series

A. T. Walden, *Senior Member, IEEE* and Zia Ziang Leong

***Abstract*—We examine Fourier transforms of real-valued stationary time series from the point of view of statistical propriety. Processes with a large dynamic range spectrum have transforms that are very significantly improper for some frequencies; the real and imaginary parts can be highly correlated, and the periodogram will not have the standard chi-square distribution at these frequencies, nor have two degrees of freedom. Use of a taper reduces impropriety to just frequencies close to zero and Nyquist only, and frequency ranges where propriety breaks down can be quite accurately and easily predicted by half the autocorrelation width of $|H * H(2f)|$, denoted by c, where $H(f)$ is the Fourier transform of the taper and $*$ denotes convolution. For vector time series we derive an improved distributional approximation for minus twice the log of the generalized likelihood ratio test (GLRT) statistic for testing for propriety of the Fourier transform at any frequency, and compare frequency range cut-offs for propriety determined by the hypothesis test with those determined by c.**

***Index Terms*—Fourier transform of time series, generalized likelihood ratio test (GLRT), improper complex vector, spectral analysis, tapering, vector-valued time series.**

I. Introduction

A complex-valued random vector $\boldsymbol{Z}$ is said to be proper, or circular to second-order, if it is uncorrelated with its complex conjugate [11], [15]. In recent years there has been a great deal of interest in the concept of improper signals, for which this second-order circularity does not hold, since performance advantages accrue by processing such signals in a way which exploits the extra information contained in the signal [2], [7], [17].

The focus of this paper is on the ubiquitous case of Fourier transforms of finite-length segments of real-valued time series, which we examine from the point of view of statistical propriety.

Consider a real-valued discrete time stationary process $\{X_t, t \in \mathbb{Z}\}$ and without loss of generality take the process to have zero mean. Let the sample interval be $\Delta_{\rm t} = 1$ so that the Nyquist frequency is $1/2$. The Fourier transform of the finite segment $\{X_t, t = 0, \ldots, N-1\}$ is given by

$$J(f) = N^{-1/2} \sum_{t=0}^{N-1} X_t \mathrm{e}^{-\mathrm{i}2\pi f t}, \ |f| \le 1/2. \qquad (1)$$

At frequencies $0 < f < 1/2$ the quantity $J(f)$ is complex-valued and the question arises as to whether it is correlated with its complex-conjugate, i.e., is proper, or not. Why should we care? For $J(f)$ Gaussian-distributed, a reasonable assumption in view of (1) and the central limit theorem, this is an important question, since if it is proper, the real and imaginary parts are independent, and consequently the periodogram $|J(f)|^2$ will have a scaled χ^2_2 distribution (chi-square with two degrees of freedom). This is a basic assumption made for the periodogram in much statistical work involving the spectrum, and yet will be invalid if $J(f)$ is improper. Indeed, the magnitude of an improper complex-valued Gaussian random variable has an 'improper/noncircular Rayleigh pdf,' [2, p. 5108].

Propriety of $J(f)$ was briefly considered in [9], (and used by [16] in the context of radar signals), where it was concluded that "Unless the noise is white, the circular anomaly will usually not be zero, and the second-order statistics must take a prescribed form that is not circular." This statement is quite vague and the purpose of the current paper is to explore the issue much further, and in particular look at the interaction between tapering and propriety. (Tapering is not mentioned in [9] or [16].) We find that impropriety is worse at frequencies where power is relatively low in comparison to any high power peaks in the spectrum, i.e., those frequencies where the spectral estimate is subject to 'side-lobe leakage.' Our results are practically useful: tapering protects against such impropriety at all but the lowest and highest frequencies, and moreover we can determine numerically what is meant by 'lowest' and 'highest,' enabling us to know where standard distributional properties will be invalid. We also compare these frequency range cut-offs for propriety with those determined by an improved generalized likelihood ratio test.

A. Contributions

Following some background in Section II on complex-valued vectors, and their statistical properties, the contributions of this paper are as follows:

1) We show that, for processes with a spectrum having a large dynamic range, the Fourier transform values are very significantly improper for frequencies affected by side-lobe leakage. The real and imaginary parts can be highly correlated and the periodogram will not have a chi-square distribution at these frequencies, nor have two degrees of freedom. This *distributional* effect complements the known problem of leakage *bias* of the periodogram in such cases.
2) The application of a taper in the Fourier transform will reduce impropriety just to frequencies close to zero and Nyquist, ensuring that the chi-square distribution for the

A. T. Walden and Z. Z. Leong are with the Dept. of Mathematics, Imperial College London, London SW7 2AZ, UK (e-mail: a.walden@imperial.ac.uk, zia.leong16@imperial.ac.uk)

periodogram is valid for most of the frequency range. Moreover, the frequencies where propriety of the Fourier transform breaks down near zero and Nyquist can be quite accurately predicted from half the autocorrelation width of $|H * H(2f)|$, denoted by c, where $H(f)$ is the Fourier transform of the taper and $*$ denotes convolution.

3) We consider making decisions on impropriety of the Fourier transform via hypothesis testing. For vector time series we derive an improved distributional approximation for minus twice the log of the generalized likelihood ratio (GLRT) statistic for testing for propriety of the Fourier transform at any frequency. This enables us to compare frequency range cut-offs for propriety determined by the hypothesis test with those determined by c; the latter is indicated by an extreme percentage point of the former.

B. Terminology

Superscripts: * denotes complex-conjugate, ^{T} denotes transpose, and ^{H} denotes complex-conjugate (Hermitian) transpose. $\mathrm{tr}\{\cdot\}$ denotes matrix trace. $\stackrel{\mathrm{d}}{\to}$ denotes convergence in distribution. $\stackrel{\mathrm{def}}{=}$ means equal by definition. $\stackrel{\mathrm{d}}{=}$ means equal in distribution. $\mathrm{Re}(\boldsymbol{Z})$ and $\mathrm{Im}(\boldsymbol{Z})$ denote the real and imaginary parts, respectively, of $\boldsymbol{Z}$.

$\mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ will denote the usual real-valued Gaussian distribution of dimension p with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathcal{N}_p^C(\mathbf{0}, \boldsymbol{\Sigma})$ will denote a *proper* complex-valued Gaussian distribution of dimension p with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. Complex-valued Random Vectors

A. Background

Let $\boldsymbol{Z} = [Z_1, \ldots, Z_p]^T$ denote a complex-valued random column vector with mean zero. The covariance matrix $\mathrm{cov}\{\boldsymbol{Z}, \boldsymbol{Z}\}$ of the complex-valued vector $\boldsymbol{Z}$ is defined as $\boldsymbol{\Sigma} = \mathrm{cov}\{\boldsymbol{Z}, \boldsymbol{Z}\} \stackrel{\mathrm{def}}{=} E\{\boldsymbol{Z}\boldsymbol{Z}^H\}$, and is Hermitian. Since it is a covariance matrix it is positive semidefinite. Additionally, $\boldsymbol{Z}$ has a *complementary covariance matrix* $\mathrm{ccov}\{\boldsymbol{Z}, \boldsymbol{Z}\}$ [17], defined as $\boldsymbol{R} = \mathrm{ccov}\{\boldsymbol{Z}, \boldsymbol{Z}\} \stackrel{\mathrm{def}}{=} \mathrm{cov}\{\boldsymbol{Z}, \boldsymbol{Z}^*\} = E\{\boldsymbol{Z}\boldsymbol{Z}^T\}$, which is complex-valued and symmetric, but not in general positive semidefinite. (The complementary covariance matrix is also called the pseudo-covariance matrix, e.g., [17].)

In the scalar case the two second central moments are the variance and the complementary variance:

$$\mathrm{var}\{Z\} = E\{|Z|^2\} \quad \text{and} \quad \mathrm{cvar}\{Z\} = E\{Z^2\}. \tag{2}$$

The two matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{R}$ both appear in the so-called *augmented covariance matrix* [17]. Suppose we form the augmented vector $\boldsymbol{U}$ by joining $\boldsymbol{Z}$ and $\boldsymbol{Z}^*$, i.e.,

$$\boldsymbol{U} = [\boldsymbol{Z}^T, \boldsymbol{Z}^H]^T = [Z_1, \ldots, Z_p, Z_1^*, \ldots, Z_p{}^*]^T, \tag{3}$$

then letting $\boldsymbol{\Sigma}_{\boldsymbol{U}} = E\{\boldsymbol{U}\boldsymbol{U}^H\}$, we obtain

$$\boldsymbol{\Sigma}_{\boldsymbol{U}} = E\left\{ \begin{bmatrix} \boldsymbol{Z} \\ \boldsymbol{Z}^* \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}^H & \boldsymbol{Z}^T \end{bmatrix} \right\} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{R} \\ \boldsymbol{R}^* & \boldsymbol{\Sigma}^* \end{bmatrix}.$$

Note that $\boldsymbol{\Sigma}_{\boldsymbol{U}}$ captures all the second-order statistics of $\boldsymbol{Z}$; $\boldsymbol{U}$ is simply a convenient structure for obtaining the matrix.

We can also write $\boldsymbol{Z}$ as $\boldsymbol{Z} = \boldsymbol{A} + \mathrm{i}\boldsymbol{B}$, where $\boldsymbol{A} = \mathrm{Re}(\boldsymbol{Z}), \boldsymbol{B} = \mathrm{Im}(\boldsymbol{Z})$. Let $\boldsymbol{\Sigma}_{\boldsymbol{V}}$ denote the covariance matrix of the real-valued random vector

$$\boldsymbol{V} = [\boldsymbol{A}^T, \boldsymbol{B}^T]^T = [A_1, \ldots, A_p, B_1, \ldots, B_p]^T,$$

and suppose we partition up the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{V}} = E\{\boldsymbol{V}\boldsymbol{V}^T\}$ into its four constituent $p \times p$ matrices,

$$\boldsymbol{\Sigma}_{\boldsymbol{V}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{AA}} & \boldsymbol{\Sigma}_{\boldsymbol{AB}} \\ \boldsymbol{\Sigma}_{\boldsymbol{BA}} & \boldsymbol{\Sigma}_{\boldsymbol{BB}} \end{bmatrix}.$$

Then

$$\boldsymbol{\Sigma}_{\boldsymbol{Z}} = E\{\boldsymbol{Z}\boldsymbol{Z}^H\} = (\boldsymbol{\Sigma}_{\boldsymbol{AA}} + \boldsymbol{\Sigma}_{\boldsymbol{BB}}) + \mathrm{i}(\boldsymbol{\Sigma}_{\boldsymbol{BA}} - \boldsymbol{\Sigma}_{\boldsymbol{AB}})$$
$$\boldsymbol{R}_{\boldsymbol{Z}} = E\{\boldsymbol{Z}\boldsymbol{Z}^T\} = (\boldsymbol{\Sigma}_{\boldsymbol{AA}} - \boldsymbol{\Sigma}_{\boldsymbol{BB}}) + \mathrm{i}(\boldsymbol{\Sigma}_{\boldsymbol{BA}} + \boldsymbol{\Sigma}_{\boldsymbol{AB}}).$$

Note that when $\boldsymbol{R}_{\boldsymbol{Z}} = \mathbf{0}$, we obtain

$$\boldsymbol{\Sigma}_{\boldsymbol{AA}} = \boldsymbol{\Sigma}_{\boldsymbol{BB}} = \tfrac{1}{2}\mathrm{Re}(\boldsymbol{\Sigma}_{\boldsymbol{Z}});\ \boldsymbol{\Sigma}_{\boldsymbol{AB}} = -\boldsymbol{\Sigma}_{\boldsymbol{BA}} = -\tfrac{1}{2}\mathrm{Im}(\boldsymbol{\Sigma}_{\boldsymbol{Z}}). \tag{4}$$

A complex-valued random vector that has $\boldsymbol{R} = \mathbf{0}$ is called a *proper* complex-valued random vector [11]. If $\boldsymbol{Z}$ is proper, it is uncorrelated with its complex conjugate $\boldsymbol{Z}^*$. When $\boldsymbol{R} \neq \mathbf{0}$, the vector is an *improper* complex-valued random vector [17].

Remark 1: From (4), for the scalar case, we have $\sigma_A^2 = \sigma_B^2 = \sigma_Z^2/2$ and $\sigma_{AB} = \sigma_{BA}$ must be zero, i.e., for a proper complex-valued random variable, the real and imaginary parts are uncorrelated with equal variances. □

A proper complex-valued random vector is also said to be '*circular* to second-order' [15]: if we multiply $\boldsymbol{Z}$ by $\mathrm{e}^{\mathrm{i}\alpha}$, for $0 \leq \alpha \leq 2\pi$, its second-order properties remain the same, i.e.,

$$E\{\boldsymbol{Z}\mathrm{e}^{\mathrm{i}\alpha}(\boldsymbol{Z}\mathrm{e}^{\mathrm{i}\alpha})^H\} = E\{\boldsymbol{Z}\boldsymbol{Z}^H\}$$
$$E\{\boldsymbol{Z}\mathrm{e}^{\mathrm{i}\alpha}(\boldsymbol{Z}\mathrm{e}^{\mathrm{i}\alpha})^T\} = \mathrm{e}^{\mathrm{i}2\alpha}E\{\boldsymbol{Z}\boldsymbol{Z}^T\} = \mathrm{e}^{\mathrm{i}2\alpha}\boldsymbol{R}_{\boldsymbol{Z}} = \mathbf{0}.$$

A random vector is (fully) circular if its probability distribution is rotationally invariant [17, p. 53].

B. Complex-Valued Gaussian Random Vector

Assume $\boldsymbol{Z} = [Z_1, \ldots, Z_p]^T$ is now a complex-valued Gaussian random vector with zero mean vector and covariance matrix $\boldsymbol{\Sigma}$. The real and imaginary parts of $\boldsymbol{Z}$ form a set of $2p$ jointly zero-mean Gaussian random variables, $\boldsymbol{V} \stackrel{\mathrm{d}}{=} \mathcal{N}_{2p}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{V}})$.

When $\boldsymbol{R} = \mathbf{0}$, $\boldsymbol{Z} \stackrel{\mathrm{d}}{=} \mathcal{N}_p^C(\mathbf{0}, \boldsymbol{\Sigma})$. Since the joint moments of orders 1 and 2 determine the entire distributional structure for the Gaussian distribution, it follows that a complex-valued zero-mean Gaussian random vector is proper iff it is (fully) circular [17, p. 53].

Remark 2: In the scalar case, if $Z \stackrel{\mathrm{d}}{=} \mathcal{N}_1^C(0, \sigma_Z^2)$, then A and B are independent, each $\mathcal{N}_1(0, \sigma_Z^2/2)$; see Remark 1. □

III. Impropriety of Scalar Fourier Transform

Under appropriate mixing conditions for a real-valued non-Gaussian stationary time series, or directly for a stationary Gaussian time series, the Fourier transform in (1) is asymptotically ($N \to \infty$) distributed as [5, p. 120],

$$J(f) \stackrel{\mathrm{d}}{=} \begin{cases} \mathcal{N}_1^C(0, S(f)) & \text{if } f \neq 0, \pm 1/2 \\ \mathcal{N}_1(0, S(f)) & \text{if } f = 0, \pm 1/2, \end{cases} \tag{5}$$

where we have concentrated on the principal domain $|f| \leq 1/2$, given that the result is periodic with period unity. So, asymptotically, the Fourier transform is proper/circular for frequencies away from zero and $\pm 1/2$. In this paper we look at these properties for finite samples.

Remark 3: Under the same conditions giving (5), we have the following distributional result for the periodogram:

$$|J(f)|^2 \stackrel{\mathrm{d}}{=} (S(f)/2)\chi^2_2, \quad f \neq 0, \pm 1/2. \tag{6}$$

□

A. Raw DFT

The orthonormal discrete Fourier transform (DFT) corresponds to evaluating the Fourier transform (1) at the Fourier frequencies $f_\ell = \ell/N$, $\ell = 0, \ldots, N-1$. We can write the DFT as $\mathcal{F}\boldsymbol{X}$, where $\boldsymbol{X}$ is a column vector with elements $\{X_t, t = 0, \ldots, N-1\}$ and $\mathcal{F}$ is an $N \times N$ matrix whose (ℓ, t)th element is $\exp(-\mathrm{i}2\pi t\ell/N)/\sqrt{N}$, for $0 \leq \ell, t \leq N-1$. A linear transform of a proper complex-valued vector is proper [11], and all scalar components are then proper. So the DFT of a proper complex-valued vector gives rise to proper complex variables at each Fourier frequency. This property was exploited in [13] for the simulation of proper scalar complex-valued Gaussian proceses. However, if $\boldsymbol{X}$ is *real-valued*, the case of interest to us, no such assurance is forthcoming.

The autocovariance sequence (ACVS) of the stationary, real-valued, zero-mean process $\{X_t\}$ is defined as

$$s_\tau \stackrel{\mathrm{def}}{=} E\{X_{t+\tau}X_t\}, \tag{7}$$

where τ is the lag. The ACVS is assumed absolutely summable so that its Fourier transform (the spectrum) exists.

Some aspects of circularity of the DFT were examined by Edelblute [9], who derived some useful analytical expressions which we now explore further.

The power or variance of the DFT is

$$P(f_\ell) = E\{|J(f_\ell)|^2\} = \frac{1}{N}\sum_{t=0}^{N-1}\sum_{u=0}^{N-1} s_{t-u}\cos\left(2\pi f_\ell(t-u)\right). \tag{8}$$

This is also the expected value of the periodogram. The covariance of the real and imaginary part of $J(f_\ell)$ is

$$C(f_\ell) = E\{\mathrm{Re}(J(f_\ell))\mathrm{Im}(J(f_\ell))\} = -\frac{1}{N}\sum_{t=1}^{N-1} s_t \sin(2\pi f_\ell t). \tag{9}$$

This is called the *circular anomaly* in [9].

The difference between the variances of the real and imaginary parts of $J(f_\ell)$ is

$$Q(f_\ell) = E\{\mathrm{Re}^2(J(f_\ell))\} - E\{\mathrm{Im}^2(J(f_\ell))\} \tag{10}$$

$$= \begin{cases} \frac{1}{N}\sum_{t=0}^{N-1}\sum_{u=0}^{N-1} s_{t-u}, & f_\ell = 0 \\ \frac{1}{N}\sum_{t=0}^{N-1}\sum_{u=0}^{N-1} s_{t-u}\cos(\pi(t+u)), & f_\ell = \pm 1/2 \\ 2\cot\left(2\pi f_\ell\right) C(f_\ell), & \text{otherwise.} \end{cases} \tag{11}$$

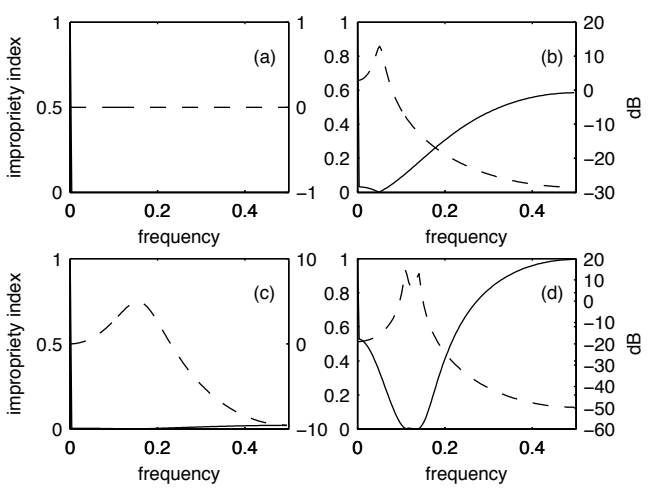

Fig. 1. Impropriety index for four time series (a)–(d) for $N = 256$. The solid line is the index, $|\varrho(f_\ell)|$, while the dashed line is the theoretical spectrum $S(f_\ell)$ measured on a decibel ($10\log_{10}$) scale.

Remark 4: The two special cases in (11) were not included in [9]. □

In addition, we can also obtain the *pseudo-variance* $R(f_\ell) = E\{J^2(f_\ell)\}$ by using Equation (9) and (11):

$$R(f_\ell) = E\{[\mathrm{Re}\,(J(f_\ell)) + \mathrm{iIm}\,(J(f_\ell))]^2\}$$
$$= Q(f_\ell) + 2\mathrm{i}C(f_\ell) \tag{12}$$

$$= \begin{cases} \frac{1}{N}\sum_{t=0}^{N-1}\sum_{u=0}^{N-1} s_{t-u}, & f_\ell = 0 \\ \frac{1}{N}\sum_{t=0}^{N-1}\sum_{u=0}^{N-1} s_{t-u}\cos(\pi(t+u)), & f_\ell = \pm 1/2 \\ 2C(f_\ell)\csc\left(2\pi f_\ell\right)\mathrm{e}^{\mathrm{i}2\pi f_\ell}, & \text{otherwise.} \end{cases} \tag{13}$$

Here csc denotes cosec.

Ollila [12] considered the ratio between the pseudo-variance and the variance of a complex random variable, Z, namely $\varrho_Z \stackrel{\mathrm{def}}{=} R_Z/P_Z$. It measures the correlation of the variable with its complex conjugate. Its modulus quantifies how improper is Z : it is proper if $|\varrho_Z| = 0$ and it is maximally improper if $|\varrho_Z| = 1$. If Z is Gaussian, then $|\varrho_Z| = 0$ corresponds to circularity, otherwise, just second-order circularity. $|\varrho_Z|$ was called the circularity coefficient in [12], but we prefer 'impropriety index' which avoids confusion over second or higher-order circularity.

For each Fourier frequency f_ℓ we can use (8) and (13) to calculate the impropriety index $|\varrho(f_\ell)| = |R(f_\ell)/P(f_\ell)|$ analytically. We have done this for four different time series models:

$$\begin{aligned} &\text{(a) } X_t = \epsilon_t \\ &\text{(b) } X_t = 1.8X_{t-1} - 0.9X_{t-2} + \epsilon_t \\ &\text{(c) } X_t = 0.75X_{t-1} - 0.5X_{t-2} + \epsilon_t \\ &\text{(d) } X_t = 2.7607X_{t-1} - 3.8106X_{t-2} \\ &\qquad + 2.6535X_{t-3} - 0.9238X_{t-4} + \epsilon_t \end{aligned} \tag{14}$$

Here $\{\epsilon_t\}$ is zero-mean Gaussian white noise. Model (b) is taken from [21]. Models (c) and (d) are taken from [14]. Models (b) and (d) have large spectrum dynamic ranges of

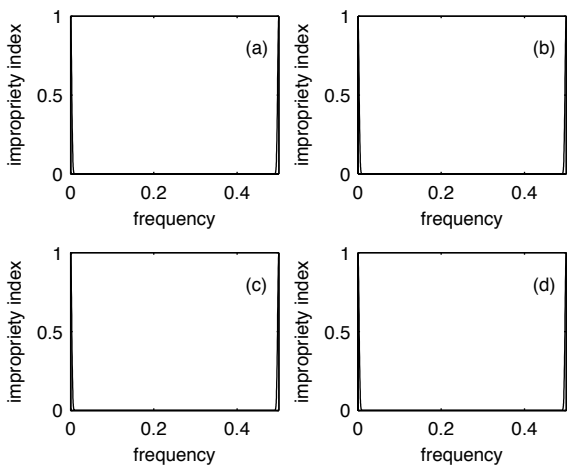

Fig. 2. Impropriety index $|\varrho_h(f'_\ell)|$ for finer frequency approach with tapering, $N = 256, N' = 4096$.

41dB and 65dB, respectively, (where the decibel (dB) scale is a $10\log 10$ scale), and their spectral magnitudes vary quite rapidly with frequency. The variance of $\{\epsilon_t\}$ is chosen to make $\{X_t\}$ have unity variance. Only model (a) was considered in [9].

For each model, and $N = 256$, Fig. 1 shows the impropriety index (solid line, left y-axis) with frequency, while the spectral density function $S(f_\ell) = \sum_{\tau=-\infty}^{\infty} s_\tau \mathrm{e}^{-\mathrm{i}2\pi f_\ell \tau}$, is also shown (dashed line, right y-axis). In spectral estimation, "side-lobe leakage" describes the effect of estimated power transferring from high power parts of the spectrum to low power parts. Such leakage is absent when the spectrum is flat (white noise) and most acute when the dynamic range is high. We see that the impropriety index is zero for white noise, apart from at frequency zero. Also for the low dynamic range spectrum of model (c) there is only a small impropriety away from the peak, just visible at high frequencies where the spectrum is lowest. However, for models (b) and (d), where the dynamic range is high, impropriety is essentially zero where the spectrum is high, but where the spectrum is low, the impropriety index can become very high.

A stochastic process runs from $-\infty$ to ∞ so our observation of $X_0, \ldots, X_{N-1}$ can be viewed as multiplying the process by a box-car function which is zero before the observations, then unity, and then zero again after the observations. Such sharp features cause ripples in the Fourier transform leading to the leakage discussed above; this can be ameliorated by applying a taper, which we discuss next.

B. DFT with tapering

Edelblute's results were derived for Fourier frequencies only and did not include tapering. We have just seen that the raw DFT produces very high impropriety for time series with high dynamic range. We now consider both tapering and a finer frequency grid. We replace (1) by

$$J_h(f) \stackrel{\text{def}}{=} \sum_{t=0}^{N-1} h_t X_t \mathrm{e}^{-\mathrm{i}2\pi f t}, \tag{15}$$

where $\{h_t,\, t = 0, \ldots, N-1\}$ is a data taper normalized to have a sum of squares of unity. Then the tapered periodogram is given by $|J_h(f)|^2|$.

We utilise the zeroth-order Slepian taper: let $H(f) = \sum_{t=0}^{N-1} h_t \mathrm{e}^{-\mathrm{i}2\pi f t}$, be the Fourier transform of the taper, and let $\mathcal{H}(f) = |H(f)|^2$ be the corresponding spectral window. Then the zeroth-order Slepian taper is the sequence such that $\mathcal{H}(f)$ maximizes the concentration ratio $[\int_{-W}^{W} \mathcal{H}(f)\mathrm{d}f]/[\int_{-1/2}^{1/2} \mathcal{H}(f)\mathrm{d}f]$ over a chosen design interval $[-W, W] \subseteq [-1/2, 1/2]$. Many further details are provided in [14, Chapters 7 and 8]. Such a taper reduces leakage through excellent sidelobe suppression in $\mathcal{H}(f)$.

The effective bandwidth, B say, associated with the taper, can be equated to the autocorrelation width [4, p. 154] of $\mathcal{H}(f) = |H(f)|^2$:

$$B \stackrel{\text{def}}{=} \mathrm{width_a}\{|H(f)|^2\} = \frac{\left\{\int_{-1/2}^{1/2} |H(f)|^2 \mathrm{d}f\right\}^2}{\int_{-1/2}^{1/2} |H(f)|^4 \mathrm{d}f}. \tag{16}$$

As shown in [22, p. 210], the effective bandwidth for the zeroth-order Slepian taper is actually less than that of the 'design bandwidth' of $2W$. In this work we chose $W = 3/N$, so that the associated effective bandwidth is then less than $2W = 6/N = 0.023$, i.e., not too wide in terms of Fourier frequencies and consistent with resolving the spectra in Fig. 1.

Let $P_h(f) \stackrel{\text{def}}{=} E\{|J_h(f)|^2\}, R_h(f) \stackrel{\text{def}}{=} E\{J_h^2(f)\}$. Then [21],

$$P_h(f) = \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} h_t h_u s_{t-u} \mathrm{e}^{-\mathrm{i}2\pi f(t-u)} \tag{17}$$

$$R_h(f) = \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} h_t h_u s_{t-u} \mathrm{e}^{-\mathrm{i}2\pi f(t+u)}. \tag{18}$$

We can pad with zeros to give a length $N' = 2^n >> N$ and thus calculate $P_h(f'_\ell)$ and $R_h(f'_\ell)$ where $f'_\ell = \ell/N'$, $\ell = 0, \ldots, N'-1$, giving a much finer frequency scale. To obtain (17) we use a forward fast Fourier transform (FFT) over t, and then an inverse FFT over u, and to calculate (18) we use two forward FFTs. With $N = 256$ we chose to use $N' = 4096$. Fig. 2 shows the impropriety index, $|\varrho_h(f'_\ell)| = |R_h(f'_\ell)/P_h(f'_\ell)|$, in this case. Away from $f = 0, 1/2$ the impropriety index is virtually zero. Clearly the tapering has been very beneficial in producing a proper or (second-order) circular outcome for most of the frequency range.

However, it is highly informative to examine behaviour very close to zero and Nyquist. Fig. 3 is a zoom-in to the low frequency end ($0 \le f'_\ell < 0.01$) of the plots in Fig. 2. The high frequency end, ($0.49 < f'_\ell \le 0.5$), is not shown, but follows analogously from the symmetry seen in Fig. 2.

It is shown in Appendix-A that $|\varrho_h(f)| = |H * H(2f)| = \left|\sum_{t=0}^{N-1} h_t^2 \mathrm{e}^{-\mathrm{i}4\pi f t}\right|$, where $*$ denotes convolution. Now $|H * H(2f)|$ is real-valued, positive and peaked about $f = 0$. So its autocorrelation width will be a suitable measure of its width:

$$\mathrm{width_a}\{|H * H(2f)|\} \stackrel{\text{def}}{=} \frac{\left\{\int_{-1/2}^{1/2} |H * H(2f)|\mathrm{d}f\right\}^2}{\int_{-1/2}^{1/2} |H * H(2f)|^2 \mathrm{d}f}. \tag{19}$$

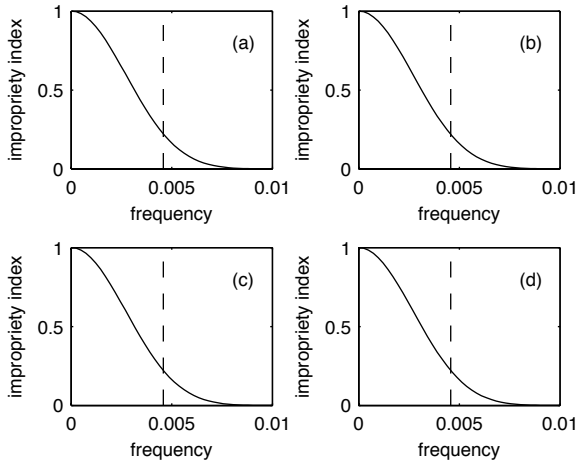

Fig. 3. Impropriety index $|\varrho(f'_\ell)|$ for finer frequency approach with tapering, $N = 256, N' = 4096$, restricted to very low frequencies. The vertical dashed line marks $c = \frac{1}{2}\mathrm{width_a}\{|H * H(2f)|\}$.

Let us denote

$$c = \tfrac{1}{2}\mathrm{width_a}\{|H * H(2f)|\}. \tag{20}$$

Impropriety is then predicted for the (positive) frequency ranges $[0, c] \cup [\frac{1}{2} - c, \frac{1}{2}]$. The quantity c is marked by a vertical dashed line in Fig. 3. This measure provides a useful estimate of the frequency extent of the impropriety under tapering.

Remark 5: Under the same conditions giving (5), we have the following distributional result for the tapered periodogram:

$$|J_h(f)|^2 \stackrel{\mathrm{d}}{=} (S(f)/2)\chi^2_2, \quad f \neq 0, \pm 1/2, \tag{21}$$

i.e., the tapered and the untapered periodograms have the same large sample distribution [5, p. 128]. □

IV. Statistical Implications of Impropriety

A. Correlation Between Real and Imaginary Parts

1) No tapering: We consider the covariance matrix of the real and imaginary parts of $J(f'_\ell)$:

$$\begin{bmatrix} E\{\mathrm{Re}^2(J(f'_\ell))\} & E\{\mathrm{Re}(J(f'_\ell))\mathrm{Im}(J(f'_\ell))\} \\ E\{\mathrm{Re}(J(f'_\ell))\mathrm{Im}(J(f'_\ell))\} & E\{\mathrm{Im}^2(J(f'_\ell))\} \end{bmatrix}.$$

Using (8) and (10), then (9) and (12), we can write this as

$$\tfrac{1}{2}\begin{bmatrix} [P(f'_\ell) + \mathrm{Re}(R(f'_\ell))] & \mathrm{Im}(R(f'_\ell)) \\ \mathrm{Im}(R(f'_\ell)) & [P(f'_\ell) - \mathrm{Re}(R(f'_\ell))] \end{bmatrix}. \tag{22}$$

The components of this matrix can be found using (17) and (18) with the taper set to the default $h_t = 1/\sqrt{N}$.

Remark 6: For $f = 0, \pm 1/2$, $J(f)$ is real-valued and the covariance matrix is singular. □

Using (22) we can find the correlation between the real and imaginary parts of $J(f'_\ell)$; under propriety, it is zero. Fig. 4 shows the correlation against frequency for our four models. For models (a), white noise, and (c), the low dynamic range example, the correlation is essentially zero apart from very close to zero and $1/2$. Now let us look at models (b) and (d). Recalling Fig. 1, we see that in parts of the frequency range where leakage is not a problem (and the impropriety index is zero) the correlation is around zero, but where there is considerable leakage (and the impropriety index is significantly non-zero) the correlation can reach high values, such as -0.87 for model (d). Correlation will mean that the scaled χ^2_2 statistical model for the periodogram — which assumes independent real and imaginary parts — will be invalid. The oscillating correlation structure seen in Figs. 4(b) and (d) will affect the degrees of freedom of the periodogram, which we examine shortly.

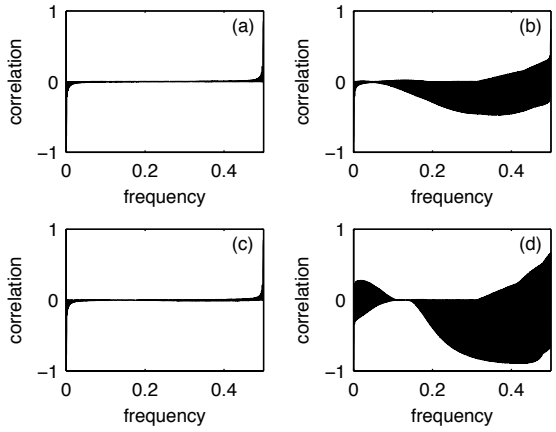

Fig. 4. Correlation between real and imaginary parts of $J(f'_\ell)$ for four time series (a)–(d). $N = 256, N' = 4096$ and no taper.

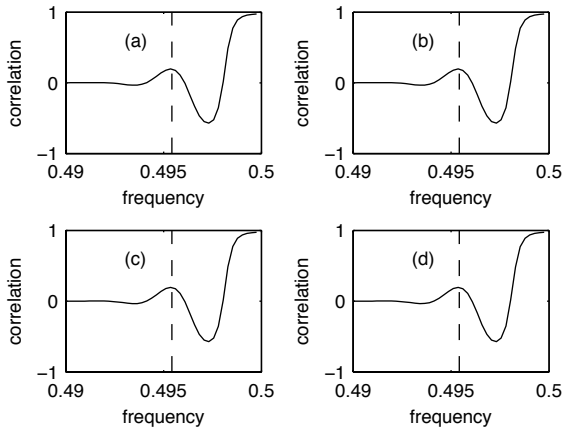

Fig. 5. Correlation between real and imaginary parts of $J_h(f'_\ell)$ for four time series (a)–(d). $N = 256, N' = 4096$ with Slepian taper, restricted to very high frequencies. The dashed line is as in Fig. 3.

2) With tapering: Just as for the impropriety index, the effect of applying a taper is dramatic: the correlation between the real and imaginary parts of $J_h(f'_\ell)$ for all four models is then only significantly non-zero close to $f = 0, \pm 1/2$. Fig. 5 shows the correlation for high frequencies $f'_\ell > 0.49$ and c gives a good indication of the frequency below which the correlation can be taken to be zero. At low frequencies the plot is inverted with the correlation falling to -1 as f'_ℓ approaches zero.

B. Distribution

The probability density function (PDF) of the magnitude of an improper complex-valued Gaussian random variable was given in [2, p. 5108], who called it 'the improper/noncircular Rayleigh pdf.' The PDF of $\hat{S}(f) \stackrel{\text{def}}{=} |J(f)|^2$ then follows as

$$g(\hat{S}(f)) = \frac{1}{P(f)\sqrt{\eta}(f)} \mathrm{e}^{-\hat{S}(f)/[P(f)\eta(f)]} \, I_0\left(\frac{\hat{S}(f)|\rho(f)|}{P(f)\eta(f)}\right), \tag{23}$$

$\hat{S}(f) > 0$, where $\eta(f) \stackrel{\text{def}}{=} 1 - |\rho(f)|^2$, and $I_0(\cdot)$ is the modified Bessel function of the first kind of order zero. Note that when $\rho(f) = 0$, the PDF is that of an exponential random variable with mean $P(f)$:

$$g(\hat{S}(f)) = \frac{1}{P(f)} \mathrm{e}^{-\hat{S}(f)/P(f)}, \quad \hat{S}(f) > 0, \tag{24}$$

or, $|J(f)|^2 \stackrel{\text{d}}{=} (P(f)/2)\chi^2_2$.

- At a frequency f_a where the periodogram is unaffected by side-lobe leakage and local bias is negligible, $E\{|J(f_a)|^2\} \stackrel{\text{def}}{=} P(f_a) = S(f_a)$, and $\rho(f_a) = 0$, so that (6) or (24) gives the standard distributional properties.
- At a frequency f_b where the periodogram is affected by side-lobe leakage, $P(f_b) > S(f_b)$, and $\rho(f_b) \neq 0$, and the PDF (23) applies.

Fig. 6 illustrates these results: here f_a is the nearest Fourier frequency to the peak in the SDF for model (b), shown in Fig. 1(b). f_b is chosen to be 0.4 for the same model. The plots in the top row of Fig. 6 show exactly the behaviour discussed above. For the plots in the bottom row tapering has been applied, side-lobe leakage is eliminated, and the distributional behaviour now also obeys standard theory at f_b.

C. Degrees of Freedom

All the model-based properties shown so far are are exact. We now look at the degrees of freedom of the periodogram via simulation. We take the periodogram to be distributed as a scaled chi-squared variable with ν degrees of freedom: $\hat{S}(f) \stackrel{\text{d}}{=} a\chi^2_\nu$, for which $E\{\hat{S}(f)\} = a\nu$ and $\mathrm{var}\{\hat{S}(f)\} = 2a^2\nu$, so that

$$\nu = 2\frac{\left(E\{\hat{S}(f)\}\right)^2}{\mathrm{var}\{\hat{S}(f)\}}. \tag{25}$$

The quantity ν is often called the equivalent degrees of freedom of the spectrum estimator $\hat{S}(f)$ [14, p. 255]. To estimate ν at each frequency we can simulate a time series from the chosen model process m times, calculate the sample means and variances of $\hat{S}(f'_\ell)$, and use these in (25) to obtain $\hat{\nu}$. This was done using $m = 10\,000$ and the results are shown in Fig. 7 for model (d). Plots (a)–(c) are for no tapering and we see that the degrees of freedom oscillate between 1.5 and 2 at the low-frequency end and between 1 and 2 at the high-frequency end.

When tapering is applied the effect is immediate: in plot (d) we see at the high-frequency end that the degrees of freedom for $\hat{P}_h(f)$ are 2 until the boundary marked by the vertical line implied by (19), after which the degrees of freedom roll-off smoothly to unity. (The same effect is observed at the low-frequency end.)

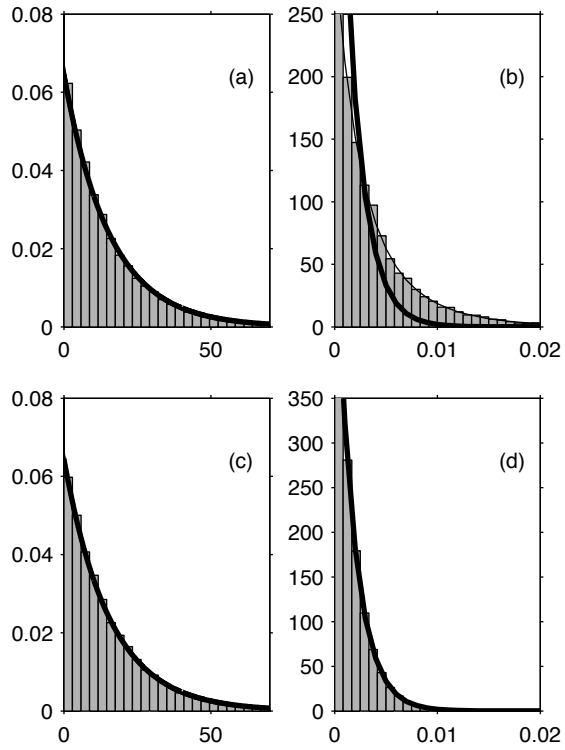

Fig. 6. Theoretical and empirical distributions of periodogram estimator for model (b), at frequency f_a, (left column) and frequency f_b, (right column), without tapering, (first row), and with tapering (second row). In every plot the thick line gives the PDF in (6), and the thin line that in (23); they are only noticeably different in plot (b). Here $N = 256$ and 20 000 simulations were used to build the histograms.

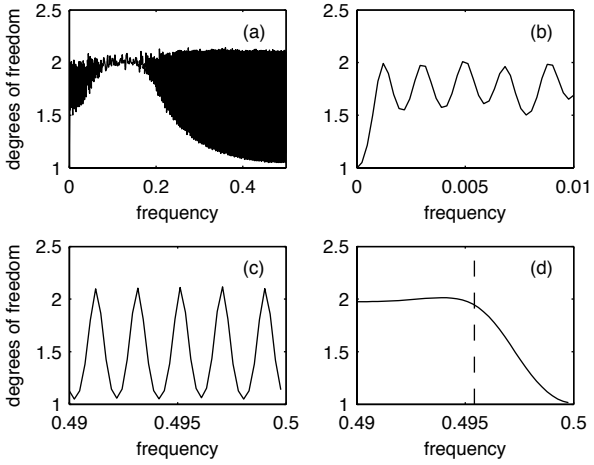

Fig. 7. Estimated degrees of freedom for model (d) using $N = 256$ and $N' = 4096$. Plots (a)–(c) are for no tapering, and show the full frequency range, the low-frequency end and the high-frequency-end, respectively. Plot (d) shows the high-frequency end when the Slepian taper is used, and the vertical dashed line is as in Fig. 3.

D. The Case for Tapering

Brillinger [6], considered tapering in spectrum estimation. His two main advantages for tapering were (i) "an improve-

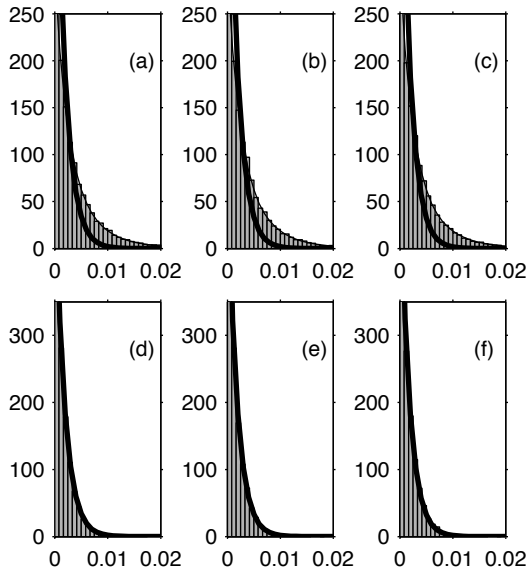

Fig. 8. All plots for model (b). Top row: theoretical and empirical distributions of periodogram, at frequencies (a) $f_b - \frac{1}{N}$, (b) f_b, (c) $f_b + \frac{1}{N}$. Bottom row: theoretical and empirical distributions of tapered periodograms, all at frequency f_b, using Slepian tapers of orders 0, 1, 2 for plots (d)–(f), respectively. In each plot the thick line gives the PDF in (6), and the thin line that in (23). Here $N = 256$ and 20 000 simulations were used to build the histograms.

ment in the resolution (reduction in the bias) of the estimate in the case that the power spectrum has a split peak or a very large peak at some location" and (ii) "one can arrange that the finite Fourier transform values at distinct frequencies be more nearly uncorrelated."

In the current paper we have shown that tapering results in $J_h(\cdot)$ being proper over all the frequency range except at very high or low frequencies, and that what is meant by "very high/low" can be numerically predicted. The propriety of J_h means that standard chi-square approximations for the (tapered) periodogram will be valid.

V. A Highly Relevant Example

We now illustrate the importance of the results by consideration of two standard types of smoothing for spectrum estimation. We firstly remind ourselves of the standard theory, then consider the effects of impropriety.

A. Smoothed periodogram

The smoothed periodogram is found by averaging $2K + 1$ adjacent periodogram ordinates at the Fourier frequencies [5]:

$$\bar{S}(f_\ell) = \frac{1}{2K+1} \sum_{k=-K}^{K} \hat{S}(f_{\ell-k}). \tag{26}$$

Standard large sample theory [19, p. 196] considers $\bar{S}(f_j)$ to be the average of $2K + 1$ approximately independent random variables each distributed as in (6), so that

$$\bar{S}(f_\ell) \stackrel{\mathrm{d}}{=} \left[\frac{S(f_\ell)}{4K+2}\right] \chi^2_{4K+2}, \quad f_\ell \neq 0, \pm 1/2, \tag{27}$$

provided that the spectrum varies little over the $2K+1$ Fourier frequencies.

B. Multitapering

The multitaper spectrum estimator is found by averaging over a set of K tapered periodograms at frequency f_ℓ:

$$\hat{S}^{(MT)}(f_\ell) = \frac{1}{K} \sum_{k=0}^{K-1} |J_{h_k}(f_\ell)|^2, \tag{28}$$

where $J_{h_k}(f) \stackrel{\mathrm{def}}{=} \sum_{t=0}^{N-1} h_{t,k} X_t \mathrm{e}^{-\mathrm{i}2\pi f t}$, and $\{h_{t,k}\, t = 0, \ldots, N-1; k = 0, \ldots, K-1\}$ are a set of K orthonormal, (e.g., Slepian), data tapers [22]. Standard large sample theory [14, p. 343] takes $\hat{S}^{(MT)}(f_\ell)$ to be the average of K approximately independent random variables each distributed as in (21), so that

$$\hat{S}^{(MT)}(f_\ell) \stackrel{\mathrm{d}}{=} \frac{S(f_j)}{2K} \chi^2_{2K}, \ f_\ell \neq 0, \pm 1/2, \tag{29}$$

provided that the spectrum varies little over the bandwidth of the widest taper, namely that of order $(K-1)$, [22].

C. Effect of Impropriety

For model (b), consider the estimator (26) with $K = 1$ and centered at the Fourier frequency $f_b = 0.4$. Figs. 8(a)-(c) show empirical histograms of the periodograms for frequencies $f_b - \frac{1}{N}, f_b, f_b + \frac{1}{N}$, respectively. The standard theory PDF (6) does not hold in all three cases, and hence the result (27) will be invalid. By way of contrast, the improper/noncircular magnitude squared PDF of (23), predicts the distribution of the periodograms very well.

Turning to the estimator (28), Figs. 8(d)-(f) show empirical histograms for the tapered periodograms used in (28) when $K = 3$, all at frequency f_b. The three tapers used were Slepian tapers of orders 0, 1, 2, respectively, [22]. The standard theory PDF (6), or equivalently (21), holds in all three cases, and hence the result (29) will be valid. The in-built tapering thus avoids the complication of improper/noncircular PDFs.

VI. Hypothesis Testing and Vector Transforms

We now turn to making decisions on impropriety based on samples of the Fourier transform. We will use hypothesis testing to determine the frequency interval corresponding to a rejection of propriety and compare this with the intervals determined by $[0, c] \cup [\frac{1}{2} - c, \frac{1}{2}]$. We will do this for both the scalar $(p = 1)$ and vector $(p > 1)$ cases, and in doing so we derive a distributional approximation for the test statistic for the vector case which improves on that of [23].

A. Background

Consider a real-valued, mean-zero, Gaussian and stationary, vector-valued times series $\{\boldsymbol{X}_t\}$ where $\boldsymbol{X}_t = [X_{1,t}, \ldots, X_{p,t}]^T$. Suppose we replace (15) by

$$\boldsymbol{J}_h(f) \stackrel{\mathrm{def}}{=} \sum_{t=0}^{N-1} h_t \boldsymbol{X}_t \mathrm{e}^{-\mathrm{i}2\pi t f}, \tag{30}$$

and define the augmented vector

$$\boldsymbol{U}(f) \stackrel{\mathrm{def}}{=} [\boldsymbol{J}_h^T(f), \boldsymbol{J}_h^H(f)]^T, \tag{31}$$

which is a complex-valued vector of length $2p$. We now temporarily suppress the dependence on f. Then letting $\mathbf{\Sigma}_{\boldsymbol{U}} = E\{\boldsymbol{U}\boldsymbol{U}^{\boldsymbol{H}}\}$, we will have

$$\mathbf{\Sigma}_{\boldsymbol{U}} = E\left\{ \begin{bmatrix} \boldsymbol{J}_h \\ \boldsymbol{J}_h^* \end{bmatrix} \begin{bmatrix} \boldsymbol{J}_h^H & \boldsymbol{J}_h^T \end{bmatrix} \right\} \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{P}_h & \boldsymbol{R}_h \\ \boldsymbol{R}_h^* & \boldsymbol{P}_h^* \end{bmatrix}. \quad (32)$$

Given K independent and identically distributed random samples $\boldsymbol{J}_{h;1}, \ldots, \boldsymbol{J}_{h;K}$ we know that the likelihood for the corresponding augmented vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K$ will be [18]

$$\begin{aligned} l &= \pi^{-pK} (\det \mathbf{\Sigma}_{\boldsymbol{U}})^{-K/2} \exp\left\{ -\tfrac{1}{2} \sum_{k=1}^{K} \boldsymbol{u}_k^H \mathbf{\Sigma}_{\boldsymbol{U}}{}^{-1} \boldsymbol{u}_k \right\} \\ &= \pi^{-pK} (\det \mathbf{\Sigma}_{\boldsymbol{U}})^{-K/2} \exp\left\{ -\tfrac{K}{2} \mathrm{tr}\left(\mathbf{\Sigma}_{\boldsymbol{U}}{}^{-1} \hat{\mathbf{\Sigma}}_{\boldsymbol{U}} \right) \right\}, \end{aligned}$$

where $\hat{\mathbf{\Sigma}}_{\boldsymbol{U}}$ is the sample augmented covariance matrix

$$\hat{\mathbf{\Sigma}}_{\boldsymbol{U}} = \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{u}_k^H \boldsymbol{u}_k = \begin{bmatrix} \hat{\boldsymbol{P}}_h & \hat{\boldsymbol{R}}_h \\ \hat{\boldsymbol{R}}_h^* & \hat{\boldsymbol{P}}_h^* \end{bmatrix}. \quad (33)$$

Our hypothesis test for propriety of $\boldsymbol{J}_h$, (at frequency f), is then $H_0 : \boldsymbol{R}_h = \boldsymbol{0}$ vs $H_1 : \boldsymbol{R}_h \neq \boldsymbol{0}$.

The generalized likelihood ratio test (GLRT) statistic is,

$$L_G = \frac{\max\limits_{\boldsymbol{R}_h \in H_0} L(\boldsymbol{J}_h)}{\max\limits_{\boldsymbol{R}_h \in H_1} L(\boldsymbol{J}_h)}, \quad (34)$$

a ratio of the likelihood with $\mathbf{\Sigma}_{\boldsymbol{U}}$ having zero off-diagonal blocks, to the likelihood with unconstrained $\mathbf{\Sigma}_{\boldsymbol{U}}$. The statistic can be rexpressed as [18]

$$T \stackrel{\text{def}}{=} L_G^{2/K} = \det\{\boldsymbol{I}_p - \hat{\boldsymbol{P}}_h^{-1} \hat{\boldsymbol{R}}_h \hat{\boldsymbol{P}}_h^{-*} \hat{\boldsymbol{R}}_h^*\} = \frac{\det\{\hat{\mathbf{\Sigma}}_{\boldsymbol{U}}\}}{\det^2\{\hat{\boldsymbol{P}}_h\}}.$$

B. Existing Distributional Approximations

By Wilks' theorem [20], if H_0 is true, then as $K \to \infty$,

$$M \stackrel{\text{def}}{=} -2 \log L_G = -K \log(T) \stackrel{\text{d}}{\to} \chi_\eta^2. \quad (35)$$

Here η is the difference between the number of free real parameters under H_0 and H_1. Now $\boldsymbol{R}_h$ is complex symmetric, and there are $\frac{1}{2}(p^2 - p)$ free complex parameters in the upper triangle (excluding the diagonal) and p free complex parameters on the diagonal. Hence, we have a total of $p^2 + p$ free real parameters, and therefore $\eta = p^2 + p$.

Remark 7: Delmas *et al*. [8] have derived *asymptotic* distributions for generalized likelihood ratios for testing for impropriety in the scalar and vector cases; they consider both independent and identically distributed observations, and independent but non-identically distributed scenarios. □

Our focus in the following is the finite-sample case. Box's finite-sample distributional approximation for M is

$$M \stackrel{\text{d}}{=} \frac{K}{K-p} \chi^2_{p^2+p}, \quad (36)$$

[23], and we reject H_0 if

$$M > \frac{K}{K-p} \chi^2_{p^2+p}(1-\alpha), \quad (37)$$

where $\chi^2_{p^2+p}(1-\alpha)$ is the $100(1-\alpha)\%$ point of the chi-square distribution with $p^2 + p$ degrees of freedom. When $p = 1$ (scalar case), $T = 1 - |\hat{\varrho}|^2$. Under the null hypothesis it is known that $|\hat{\varrho}|^2$, which is a coherence estimator, has a $\mathrm{beta}(1, K-1)$ distribution, from which it follows that Box's approximation is actually exact for $p = 1$.

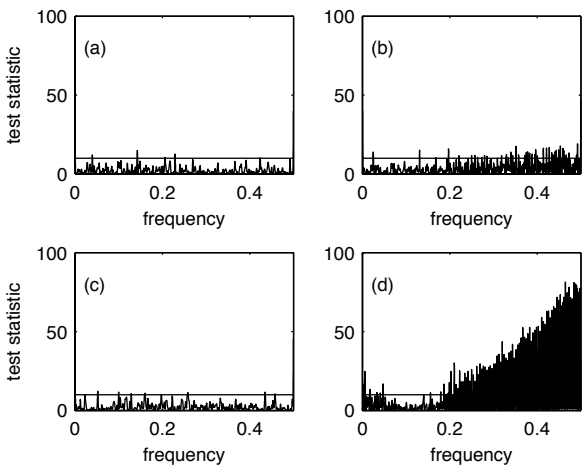

Fig. 9. GLRT test statistic $M(f'_\ell)$ for the four scalar time series (a)–(d) for $N = 256$, $N' = 4096$ and $K = 15$. No tapering used. The horizontal line is the critical value for $\alpha = 0.01$ based on Box's approximation which is exact for the scalar case.

The GLRT statistics $M(f'_\ell)$ for the four scalar time series (a)–(d) are shown in Fig. 9 for $h_t = 1/\sqrt{N}$, i.e., there is no tapering. It resulted from $K = 15$ independent simulations of length $N = 256$. The critical value for $\alpha = 0.01$ is shown by the horizontal line, and is valid at any particular frequency f'_ℓ. Viewed over all frequencies we would expect a number of false positives, but the general result is that propriety is violated for model (b) at higher frequencies, and for model (d) at most frequencies, as expected from Fig. 1.

Fig. 9 results from a one-off simulation of the GLRT; to investigate the behaviour at frequencies near zero and $1/2$ we can use averaging to get stable results. Consider replicating the test statistic L times to get $M_1(f'_\ell), \ldots, M_L(f'_\ell)$. Now each $M_i(f'_\ell)$ has the chi-square distribution with two degrees of freedom, or exponential distribution with parameter $1/2$, scaled by $K/(K-1)$. Since the sum of L independent exponential random variables Y_i has a gamma distribution, it follows that the probability density function of $\overline{Y}$, the mean of the Y's, is $g(y) = (L/2)^L [y^{L-1}/(L-1)!] \mathrm{e}^{-Ly/2}, \quad y > 0$, which identifies the $\mathrm{Gamma}(L, L/2)$ distribution. So for the mean of the M's,

$$\overline{M}(f'_\ell) \stackrel{\text{d}}{=} \frac{K}{K-1} \mathrm{Gamma}(L, L/2), \quad (38)$$

which is exact for the scalar case. Using tapering, the averaged GLRT statistic for $L = 200$ for the low-frequency end is shown in Fig. 10 for the four scalar time series. The 1% level cuts the averaged GLRT at a frequency close to c. (If the 5% level is used the match is not much different, since the 1% and 5% points of the distribution are close.) Analogous behaviour is found close to Nyquist frequency.

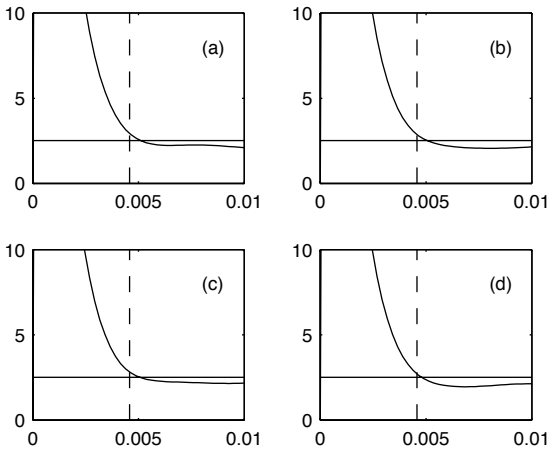

Fig. 10. Averaged GLRT statistic $\overline{M}(f'_\ell)$ for the four scalar time series (a)–(d) for $N = 256, N' = 4096, K = 15$ and $L = 200$. Low-frequency end when the Slepian taper is used. The horizontal line is the critical value for $\alpha = 0.01$ using (38) and the vertical lines are as in Fig. 3.

C. A New Distributional Approximation for Vector Case

The derivation of Box's approximation is by the idea of matching the cumulants of a scaled chi-square approximation to M up to an error of order $O(K^{-2})$. This will be problematic for $p > 1$ [3, p. 329] and leads to the consideration of other approaches.

Box [3] suggested *approximately* matching the cumulants of distributions of the form bF_{ν_1,ν_2}, i.e., a scaled F distribution with parameters ν_1, ν_2. In fact we can match the first three cumulants *exactly* as we now show, starting with Theorem 1.

Theorem 1: The $2r$th moment of $L_G(f)$, i.e., $E\{L_G^{2r}(f)\}$, is given by

$$C \prod_{j=1}^{p} \frac{\Gamma(K[1+2r]-2j+1)}{\Gamma(K[1+2r]-j+1)}, \tag{39}$$

where C is a term that does not depend on r.

Proof: This is given in AppendixB. ∎

The moment generating function (MGF) for $M(f) = -2\log L_G(f)$ is given by (with f suppressed), $\phi_M(s) = E\{\mathrm{e}^{sM}\} = E\{L_G^{-2s}\}$ so using (39),

$$\phi_M(s) = C_0 \prod_{j=1}^{p} \frac{\Gamma(K[1-2s]-2j+1)}{\Gamma(K[1-2s]-j+1)}.$$

The Gamma functions will be valid if $-2Ks+K-2j+1 > 0$ for all $j = 1, \ldots, p$, which requires $-2s > (2p-1-K)/K$.

The cumulants κ_i of M are found from the cumulant generating function by successively differentiating $\log \phi_M(s)$ and setting $s = 0$. Notice that the requirement $-2s > (2p-1-K)/K$ corresponds to $K \geq 2p$ when $s = 0$. Then, for $i \geq 1$, $\kappa_i = \mathrm{d}^i \log \phi_M(s)/(\mathrm{d}s)^i\big|_{s=0}$ so that κ_i is

$$[-2K]^i \sum_{j=1}^{p} \Big[\psi^{(i-1)}(K-2j+1) - \psi^{(i-1)}(K-j+1)\Big]. \tag{40}$$

Here for $i = 1$, $\psi(x) = [\mathrm{d}\log\Gamma(x)]/\mathrm{d}x$ is the digamma function, while for $i = 2$ and 3, $\psi^{(1)}(x)$ and $\psi^{(2)}(x)$ are the trigamma and tetragamma functions respectively; these are all 'polygamma functions.' κ_1 is the mean, κ_2 is the variance, $\kappa_3/\kappa_2^{3/2}$ is the skewness and κ_4/κ_2^2 is the excess kurtosis.

(p, K)	Method	$\alpha = 0.05$	$\alpha = 0.01$
$(2, 6)$	Asymptotic	12.59	16.81
	Box	18.89	25.22
	scaled F	19.97	26.97
	simulated	19.89	26.50
$(3, 8)$	Asymptotic	21.03	26.22
	Box	33.64	41.95
	scaled F	36.98	46.83
	simulated	36.82	47.94
$(4, 10)$	Asymptotic	31.41	37.57
	Box	52.35	62.61
	scaled F	59.17	71.98
	simulated	59.69	71.73
$(5, 12)$	Asymptotic	43.77	50.89
	Box	75.04	87.24
	scaled F	86.61	102.46
	simulated	86.23	102.41
$(5, 20)$	Asymptotic	43.77	50.89
	Box	58.36	67.86
	scaled F	59.88	69.80
	simulated	59.23	69.69

TABLE I
COMPARISON OF $100(1-\alpha)\%$ PERCENTAGE POINTS OF $M(f)$ ACCORDING TO THE ASYMPTOTIC RESULT (35), BOX'S APPROXIMATION (36), THE SCALED F METHOD (42), AND THE SIMULATION RESULT

The parameters of bF_{ν_1,ν_2} are given by the cumulants [10]

$$\begin{aligned} b &= \frac{2\kappa_1\left(\kappa_1^2\kappa_2 - \kappa_2^2 + \kappa_1\kappa_3\right)}{2\kappa_1^2\kappa_2 - 4\kappa_2^2 + 3\kappa_1\kappa_3}, \\ \nu_1 &= \frac{4\kappa_1\left(\kappa_1^2\kappa_2 - \kappa_2^2 + \kappa_1\kappa_3\right)}{4\kappa_1\kappa_2^2 - \kappa_1^2\kappa_3 + \kappa_2\kappa_3}, \\ \nu_2 &= \frac{4\kappa_1^2\kappa_2 - 8\kappa_2^2 + 6\kappa_1\kappa_3}{\kappa_1\kappa_3 - 2\kappa_2^2}. \end{aligned} \tag{41}$$

Inserting the first three cumulants specified by (40) into (41) means that the resulting parameter values b, ν_1, ν_2 specify an F distribution whose cumulants exactly match those of $M(f)$.

Then to carry out the test $M(f)$ would be compared to

$$bF_{\nu_1,\nu_2}(1-\alpha), \tag{42}$$

where $F_{\nu_1,\nu_2}(1-\alpha)$ is the $100(1-\alpha)\%$ point of the F distribution with parameters b, ν_1, ν_2 given by (41).

Table I compares the percentage points of the distribution of $M(f)$ for the asymptotic approximation (35), Box's approximation (36) and the scaled F method (42) for $\alpha = 0.05$ and 0.01. We see that Box's approximation and the F method give significantly different results from the asymptotic result for these finite sample sizes of $K = 6, 8, 10, 12, 20$. We also see, as expected, that Box's and the F methods converge for 'large' $K = 20$.

The previous approach for scalar time series can be extended to investigate impropriety for the Fourier transform of real-valued vector time series. To do so we introduce four vector autoregressive processes of unity order for dimension 2, 3, 4 and 5. These take the form

$$\boldsymbol{X}_t = \boldsymbol{\phi}^{[p]}\boldsymbol{X}_{t-1} + \boldsymbol{\epsilon}_t \tag{43}$$

where $\{\boldsymbol{\epsilon}_t\}$ is white with an identity covariance matrix and the parameter matrices chosen are

$$\boldsymbol{\phi}^{[2]} = \begin{bmatrix} 0.5 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$$

$$\boldsymbol{\phi}^{[3]} = \begin{bmatrix} -0.7003 & 0.3192 & 0.0372 \\ 0.9459 & 0.2980 & 0.6007 \\ -0.0924 & -0.1352 & 0.6506 \end{bmatrix}$$

$$\boldsymbol{\phi}^{[4]} = \begin{bmatrix} -0.3317 & 0.3975 & -0.6044 & -0.9389 \\ 0.4881 & 0 & -0.0402 & 0.8094 \\ 0.2197 & 0.2353 & 0.7189 & 0.6110 \\ 0.1534 & -0.6342 & -0.5201 & 0.7730 \end{bmatrix}$$

$$\boldsymbol{\phi}^{[5]} = \begin{bmatrix} 0.2 & 0 & -0.1 & 0 & 0.5 \\ 0.4 & -0.2 & 0 & 0.2 & 0 \\ -0.2 & 0 & 0.3 & 0 & 0.1 \\ 0.3 & 0.1 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.5 & 0.2 \end{bmatrix},$$

all giving rise to stationary time series. To carry out the GLRT we will use bF_{ν_1,ν_2} for the distribution of $M(f'_\ell)$, as justified in Section VI-B. We apply the Slepian taper, and again pay attention to the ends of the frequency range, and use averaging to get stable results. For the F approximation the choice of K requires $K \geq 2p$, and we used $K = 15$. For the moment generating function for the average of $M_1(f'_\ell), \ldots, M_L(f'_\ell)$,

$$\begin{aligned} \phi_{\overline{M}}(s) &= \phi_{(1/L)\Sigma_{i=1}^L M_i}(s) = E\left[\mathrm{e}^{s\Sigma_{i=1}^L M_i/L}\right] \\ &= \left[E\{\mathrm{e}^{s(M/L)}\}\right]^L = [\phi_{M/L}(s)]^L, \end{aligned} \tag{44}$$

since the M_i are independent. The MGF for M/L is $\phi_{M/L}(s) = E\{\mathrm{e}^{s(M/L)}\} = E\{L_G^{-2s/L}\}$, i.e.,

$$\phi_{M/L}(s) = C_0 \prod_{j=1}^{p} \frac{\Gamma(K[1-2\frac{s}{L}] - 2j + 1)}{\Gamma(K[1-2\frac{s}{L}] - j + 1)}. \tag{45}$$

The cumulant generating function is then,

$$\log \phi_{\overline{M}}(s) = L \log \phi_{M/L}(s). \tag{46}$$

Then, for $i \geq 1$, the cumulants, $\kappa_i^\star$ say, are $\kappa_i^\star = \mathrm{d}^i \log \phi_{\overline{M}}(s)/(\mathrm{d}s)^i\big|_{s=0}$ so that $\kappa_i^\star$ is

$$L\left[-\frac{2K}{L}\right]^i \sum_{j=1}^{p} \left[\psi^{(i-1)}(K-2j+1) - \psi^{(i-1)}(K-j+1)\right]. \tag{47}$$

Using these cumulants and (41) we can calculate the corresponding parameters $b^\star$, $\nu_1^\star$ and $\nu_2^\star$, say, and we take

$$\overline{M} \stackrel{\mathrm{d}}{=} b^\star F_{\nu_1^\star,\nu_2^\star}. \tag{48}$$

The averaged GLRT statistic for $L = 200$ for the low-frequency end is shown in Fig. 11 for the four vector time series with tapering. As for the scalar case, the 1% level cuts the averaged GLRT at a frequency close to c. Again, analogous behaviour is found close to Nyquist frequency. The particular extreme percent point will vary with K.

For these four models, using tapering, we also simulated the $100(1-\alpha)\%$ percentage points of $M(f)$ at $f = 0.25$, (where H_0 will hold); this was done by calculating $M(0.25)$ over 5000 independent trials. The results are shown under 'simulated' in Table I, and we see that the scaled F approximation provides a good match.

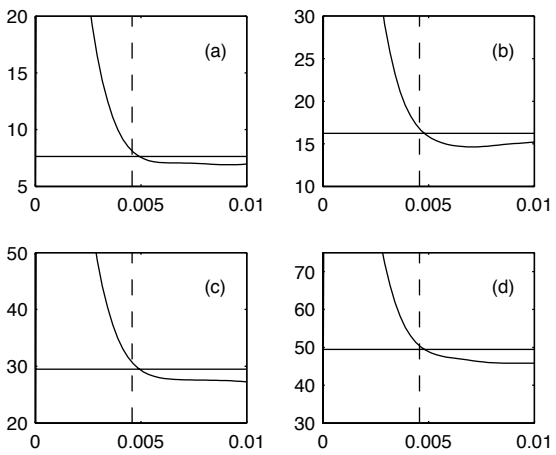

Fig. 11. Averaged GLRT statistic $\overline{M}(f'_\ell)$ for the four VAR time series (a)–(d) for $N = 256, N' = 4096, K = 15$ and $L = 200$. Low-frequency end when the Slepian taper is used. The horizontal line is the critical value for $\alpha = 0.01$ using (48) and the vertical lines are as in Fig. 3.

VII. A Useful Result

Consider the population version of the statistic T, which takes the form $\det\{\boldsymbol{I}_p - \boldsymbol{P}_h^{-1}\boldsymbol{R}_h\boldsymbol{P}_h^{-*}\boldsymbol{R}_h^*\}$. Using (54), for a designed taper this can be written as

$$\begin{aligned} &\det\{\boldsymbol{I}_p - \boldsymbol{S}^{-1}\boldsymbol{S}[H * H(2f)]\boldsymbol{S}^{-1}\boldsymbol{S}[H * H(2f)]^*\} \\ &= \det\{\boldsymbol{I}_p(1 - |H * H(2f)|^2)\} = (1 - |H * H(2f)|^2)^p. \end{aligned}$$

In the scalar case this is simply $1 - |\varrho_h(f)|^2$; see (55). We thus see that when tapering is used $|H * H(2f)|$ is a basic measure for impropriety for both the scalar and vector Fourier transforms. By design, $|H * H(2f)|$ is bounded by $\sum h_t^2 = 1$.

VIII. Summary and Conclusion

For the Fourier transform of a real-valued time series impropriety is worse at frequencies where the spectral estimate is subject to side-lobe leakage. After tapering, impropriety survives only at the lowest and highest frequencies, $[0, c] \cup [\frac{1}{2} - c, \frac{1}{2}]$, and c can be determined numerically. Hence we will know where standard distributional properties will be valid or invalid.

An improved GLRT for complex vector impropriety was derived by improving the small-sample approximation of the test statistic. This was used to compare the frequency range cut-offs for propriety determined by c, with those determined by the GLRT. The former is indicated by an extreme percentage point of the latter.

Overall, our work provides further motivation for applying tapering if carrying out statistical work with the Fourier transform of a real-valued time series.

APPENDIX

A. Impropriety and the Tapered Fourier Transform

We write the zero-mean process $\boldsymbol{X}_t$ in its spectral representation, $\boldsymbol{X}_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi\phi t}\,\mathrm{d}\boldsymbol{Z}(\phi)$, where $E\{\mathrm{d}\boldsymbol{Z}(f')\mathrm{d}\boldsymbol{Z}^H(f)\} = \boldsymbol{S}(f)\mathrm{d}f$ if $f = f'$ and zero otherwise. Here $\boldsymbol{S}(f)$ denotes the spectral matrix at frequency f. Then

$$\boldsymbol{J}_h(f) = \int_{-1/2}^{1/2} H(f-\phi)\,\mathrm{d}\boldsymbol{Z}(\phi). \tag{49}$$

Now let $J_{h;l}(f)$ denote the lth component of $\boldsymbol{J}_h(f)$, i.e., the tapered Fourier transform for series l. Using (49), with $S_{lm}(f)$ the (l,m)th element of $\boldsymbol{S}(f)$,

$$\begin{aligned} E\{J_{h;l}(f)J^*_{h;m}(f)\} &= \int_{-1/2}^{1/2} H(f-\phi)H^*(f-\phi)S_{lm}(\phi)\mathrm{d}\phi \\ &\approx S_{lm}(f)\int_{-1/2}^{1/2}|H(f-\phi)|^2\,\mathrm{d}\phi = S_{lm}(f). \end{aligned}$$

The approximation is based on the assumption that $S_{lm}(f)$ is approximately constant over the effective bandwidth, B, of $|H(f)|^2$. (This is a sensible assumption for a designed taper, but, for a spectrum with some dynamic range it would be invalid in the case of a default (no) taper $h_t = 1/\sqrt{N}$, since in that case $|H(f)|^2$ would correspond to Fejér's kernel with its extensive high sidelobes; e.g., see [14, p. 207]). The final equality uses the unit periodicity of $H(\cdot)$ and that the taper has $\sum h_t^2 = 1$. Next, $E\{J_{h;l}(f)J_{h;m}(f)\}$ is given by

$$\begin{aligned} &\int\int_{-1/2}^{1/2} H(f-\phi)H(f-\psi)E\{\mathrm{d}Z_l(\phi)\mathrm{d}Z_m(\psi)\} \\ &= \int\int_{-1/2}^{1/2} H(f-\phi)H(f-\psi)E\{\mathrm{d}Z_l(\phi)\mathrm{d}Z^*_m(-\psi)\} \\ &= \int_{-1/2}^{1/2} H(f-\phi)H(f+\phi)S_{lm}(\phi)\mathrm{d}\phi. \end{aligned}$$

The product $H(f-\phi)H(f+\phi)$ only has a significant amplitude for f close to zero and Nyquist $(1/2)$. What is meant by "f close to zero" idea is illustrated in Fig. 12, which shows from left to right the magnitude $|H(f-\phi)H(f+\phi)| = |H^*(\phi-f)H(\phi+f)|$ for "offsets" of $f = 0, B/2, B$ and $3B/2$ as a function of frequency ϕ. (Only positive frequencies are shown because of the symmetry about zero.) B is calculated using (16). Half this width, $B/2$, is also marked by a vertical dashed line in each plot. So we see that once the offset f exceeds B the magnitude of $H(f-\phi)H(f+\phi)$ is very small, (quantifying what is meant by "close to zero"), and also the magnitude is only ever large over a frequency width of B. So, assuming as before, that $S_{lm}(f)$ is approximately constant over a bandwidth B, we can take $E\{J_{h;l}(f)J_{h;m}(f)\} \approx S_{lm}(f)\int_{-1/2}^{1/2} H(f-\phi)H(f+\phi)\,\mathrm{d}\phi$, where we realize from the foregoing that only for a small range of frequencies f will this be non-zero.

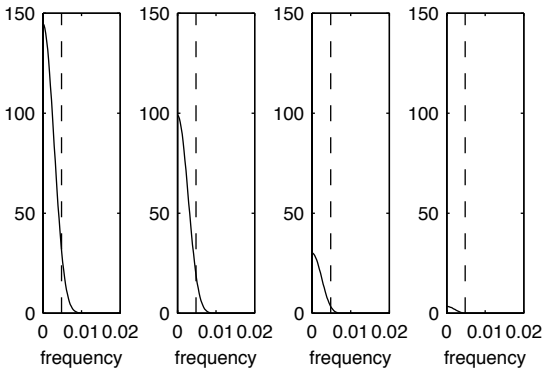

Fig. 12. Left to right: the magnitude $|H(f-\phi)H(f+\phi)| = |H^*(\phi-f)H(\phi+f)|$ for "offsets" of $f = 0, B/2, B$ and $3B/2$ as a function of frequency ϕ. $B/2$ is marked by a vertical dashed line in each plot. The Slepian taper is used here.

Now, since $H(\cdot)$ has unit periodicity,

$$\begin{aligned} \int_{-1/2}^{1/2} H(f-\phi)H(f+\phi)\,\mathrm{d}\phi &= \int_{-1/2}^{1/2} H(\psi)H(2f-\psi)\,\mathrm{d}\psi \\ &= H * H(2f) \qquad (50) \\ &= \sum_{t=0}^{N-1} h_t^2 \mathrm{e}^{-\mathrm{i}4\pi f t}. \qquad (51) \end{aligned}$$

Now $\boldsymbol{J}_h(f) = [J_{h;1}(f), \ldots, J_{h;p}(f)]^T$ so from (32),

$$\boldsymbol{P}_h(f) = \boldsymbol{S}(f), \tag{52}$$

$$\boldsymbol{R}_h(f) = \boldsymbol{S}(f)\cdot(H * H(2f)), \tag{53}$$

and therefore,

$$\boldsymbol{\Sigma}_{\boldsymbol{U}} = \boldsymbol{S}(f)\begin{bmatrix} \boldsymbol{I}_p & H*H(2f)\boldsymbol{I}_p \\ (H*H(2f))^*\boldsymbol{I}_p & \boldsymbol{I}_p \end{bmatrix}. \tag{54}$$

In the case of a scalar series, $p = 1$, (52) and (53) give

$$|\varrho_h(f)| = \frac{|R_h(f)|}{|P_h(f)|} = \frac{E\{J_h^2(f)\}}{E\{|J_h(f)|^2\}} = |H*H(2f)|. \tag{55}$$

The above results are illustrated in Fig. 13 for model (b) of (14). Here plot (b) shows that the two sides of (55) are indistinguishable for a designed taper, whereas plot (a) shows that the result does not hold for the default (no) taper $h_t = 1/\sqrt{N}$, as stated above.

B. Proof of Theorem 1

Under the null, $E\{T^r\}$ may be written as [23, p. 829]

$$2^{2rp}\frac{\prod_{j=1}^{p}\Gamma(K-j+1)\prod_{j=1}^{2p}\Gamma((2r+K-j+1)/2)}{\prod_{j=1}^{2p}\Gamma((K-j+1)/2)\prod_{j=1}^{p}\Gamma(2r+K-j+1)}. \tag{56}$$

where $\Gamma(\cdot)$ is the gamma function. We rearrange this:

$$2^{2rp}\frac{\prod_{j=1}^{p}\Gamma(K-j+1)}{\prod_{j=1}^{p}\Gamma(2r+K-j+1)}\prod_{j=1}^{2p}\frac{\Gamma((2r+K-j+1)/2)}{\Gamma((K-j+1)/2)}.$$

To simplify the last product we use the duplication formula for gamma functions [1, (6.1.18)], $\Gamma(z)\Gamma(z+\frac{1}{2}) =$

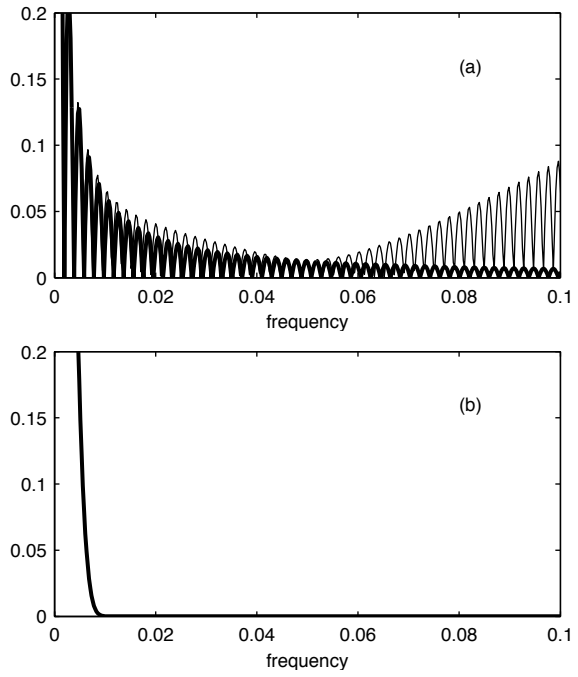

Fig. 13. Comparison of $|\varrho_h(f)|$ (solid line) versus $|H * H(2f)|$ (heavy line) for model (b) where in plot (a) the default taper $h_t = 1/\sqrt{N}$ is used, while in plot (b) the Slepian taper is used. Ranges have been limited to show any differences more clearly; for plot (b) there are no discernible differences.

$\pi^{1/2}2^{1-2z}\Gamma(2z)$. Using this we get, for $\ell = 1, \ldots, p$,

$$\Gamma((2r+K-2\ell+1)/2)\Gamma((2r+K-2\ell+2)/2) = \pi^{1/2}2^{-2r-K+2\ell}\Gamma(2r+K-2\ell+1);$$
$$\Gamma((K-2\ell+1)/2)\Gamma((K-2\ell+2)/2) = \pi^{1/2}2^{-K+2\ell}\Gamma(K-2\ell+1).$$

So $E\{T^r\}$ may now be written as

$$\frac{\prod_{j=1}^{p}\Gamma(K-j+1)}{\prod_{j=1}^{p}\Gamma(2r+K-j+1)}\prod_{j=1}^{p}\frac{\Gamma(2r+K-2j+1)}{\Gamma(K-2j+1)},$$

or, as

$$C\prod_{j=1}^{p}\frac{\Gamma(2r+K-2j+1)}{\Gamma(2r+K-j+1)}, \tag{57}$$

where C does not depend on r. Now $T^r(f) = L_G^{2r/K}(f)$ and letting $r \to rK$, gives $T^{rK}(f) = L_G^{2r}(f)$. So $E\{L_G^{2r}(f)\}$ is found by replacing r by rK in (57), which gives (39).

ACKNOWLEDGEMENT

The authors thank the referees for many helpful comments.

REFERENCES

[1] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1965.
[2] T. Adalı, P. J. Schreier & L. L. Scharf, "Complex-valued signal processing: the proper way to deal with impropriety," *IEEE Transactions on Signal Processing*, vol. 59, pp. 5101–5125, 2011.
[3] G. E. P. Box, "A general distribution theory for a class of likelihood criteria," *Biometrika*, vol. 36, 317–46, 1949.
[4] R. N. Bracewell, *The Fourier Transform and Its Applications (Second Edition)*. New York: McGraw-Hill, 1978.
[5] D. R. Brillinger, *Time Series: Data Analysis and Theory (Expanded Edition)*. New York: McGraw-Hill Inc., 1981.
[6] D. R. Brillinger, "The key role of tapering in spectrum estimation," *IEEE Trans. Acoust. Speech Sig. Proc.,* vol. 29, pp. 1075–1076, 1981.
[7] P. Chevalier, J.-P. Delmas and A. Oukaci, "Properties, performance and practical interest of the widely linear MMSE beamformer for rectilinear signals," *Signal Processing*, vol. 97, pp. 269–281, 2014.
[8] J.-P. Delmas, A. Oukaci & P. Chevalier, "On the asymptotic distribution of GLR for impropriety of complex signals," *Signal Processing,* vol. 91, pp. 2259–2267, 2011.
[9] D. J. Edelblute, "Noncircularity," *IEEE Signal Processing Letters,* vol. 3, pp. 156–157, 1996.
[10] P. Ginzberg, "Quaternion matrices: statistical properties and applications to signal processing and wavelets," Ph. D. dissertation, Dept. Mathematics, Imperial College London, 2013.
[11] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Information Theory*, vol. 39, pp. 1293–1302, 1993.
[12] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Processing Letters,* vol. 15, pp. 841–844, 2008.
[13] D. B. Percival, "Exact simulation of complex-valued Gaussian stationary processes via circulant embedding," *Signal Process.*, vol. 86, pp. 14701476, 2006.
[14] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*. Cambridge, UK: Cambridge University Press, 1993.
[15] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Transactions on Signal Processing* vol. 44, pp. 2637–2640, 1996.
[16] R. Rytel-Andrianik, "On the circularity of complex-valued radar signal," Signal Processing Symposium, Serock, Poland, June 5-7th, 2013.
[17] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data*, Cambridge UK: Cambridge University Press, 2010.
[18] P. J. Schreier, L. L. Scharf and A. Hanssen, "A generalized likelihood ratio test for impropriety of complex signals," *IEEE Signal Process. Lett.*, vol. 13, pp. 433–6, 2006.
[19] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications, 3rd Ed*, New York: Springer, 2011.
[20] G. A. Young and R. L.Smith, *Essentials of Statistical Inference*. Cambridge UK: Cambridge University Press, 2005.
[21] A. T. Walden, E. J. McCoy and D. B. Percival, "The variance of multitaper spectrum estimates for real Gaussian processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 479–482, 1994.
[22] A. T. Walden, E. J. McCoy and D. B. Percival, "The effective bandwidth of a multitaper spectral estimator," *Biometrika*, vol. 82, 201–214, 1995.
[23] A. T. Walden and P. Rubin-Delanchy, "On testing for impropriety of complex-valued Gaussian vectors," *IEEE Transactions on Signal Processing* vol. 57, pp. 825–834, 2009.