Bias correction of quadratic spectral estimators

BY LACHLAN C. ASTFALCK

School of Physics, Mathematics & Computing
The University of Western Australia, 35 Stirling Hwy, Crawley, 6009, Australia
lachlan.astfalck@uwa.edu.au

ADAM M. SYKULSKI

Department of Mathematics
Imperial College London, South Kensington, London SW7 2AZ, U.K.
adam.sykulski@imperial.ac.uk

EDWARD J. CRIPPS

School of Physics, Mathematics & Computing
The University of Western Australia, 35 Stirling Hwy, Crawley, 6009, Australia
edward.cripps@uwa.edu.au

SUMMARY

The three cardinal, statistically consistent, families of non-parametric estimators to the power spectral density of a time series are lag-window, multitaper and Welch estimators. However, when estimating power spectral densities from a finite sample each can be subject to non-ignorable bias. Astfalck et al. (2024) developed a method that offers significant bias reduction for finite samples for Welch's estimator, which this article extends to the larger family of quadratic estimators, thus offering similar theory for bias correction of lag-window and multitaper estimators as well as combinations thereof. Importantly, this theory may be used in conjunction with any and all tapers and lag-sequences designed for bias reduction, and so should be seen as an extension to valuable work in these fields, rather than a supplanting methodology. The order of computation is larger than $\mathcal{O}(n \log n)$ typical in spectral analyses, but not insurmountable in practice. Simulation studies support the theory with comparisons across variations of quadratic estimators.

Some key words: bias correction, spectral estimation, nonparametric estimation

1. QUADRATIC SPECTRAL ESTIMATORS

Denote by $\{\mathrm{X}_t\}$ a stationary zero-mean real-valued stochastic process, discretely observed at interval Δ and indexed by $t \in \mathcal{Z}$. Without loss of generality, here, and in what follows, we assume $\Delta = 1$. Assume the power spectral density $f(\omega)$ exists, and so for auto-covariance sequence $\gamma(\tau) = \mathrm{E}[\mathrm{X}_t \mathrm{X}_{t-\tau}]$,

$$\gamma(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) \mathrm{e}^{i\omega\tau} \, \mathrm{d}\omega, \quad f(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \mathrm{e}^{-i\omega\tau}$$

with $\tau \in \mathcal{Z}$ and $\omega \in [-\pi, \pi]$ is defined in radians. We make the following assumption on $\{\mathrm{X}_t\}$.

Assumption 1. Let $\{\mathrm{X}_t\}$ be a general linear process so that $\mathrm{X}_t = \sum_{i=-\infty}^{\infty} \theta_i \epsilon_{t-i}$ with independent and identically distributed ϵ_t with $\mathrm{E}[\epsilon_t] = 0$, $\mathrm{E}[\epsilon_t^2] = 1$, $\mathrm{E}[\epsilon_t^4] < \infty$, and the regularity condition $\sum_{\tau=-\infty}^{\infty} (1 + |\tau|)|\gamma(\tau)| < \infty$, implying $f(\omega)$ has a bounded and uniformly continuous derivative.

The predominant task in spectral density estimation is to obtain some estimate of the power spectral density; herein, we concern ourselves with the case of non-parametric estimation of $f(\omega)$. The periodogram is the fundamental estimator of the power spectral density, but is statistically inconsistent. Broadly, consistency is resolved by one of three families of non-parametric estimators: lag-windows, multitapers and Welch estimators. Delineating all non-parametric estimators into one of these three methodologies is not always obvious; for instance, Bartlett's estimator may be seen both as a special case of Welch's estimator or as a lag-window estimator with a triangular lag-sequence. This literature is too vast to adequately review here, and we defer the reader to one of the many comprehensive book-length treatments such as Percival & Walden (2020) and Politis & McElroy (2019) for more detailed information. We find these to be the categorisations and naming conventions most common in the present literature, but acknowledge that alternatives exist (for example, see pp.252 Percival & Walden, 2020).

Quadratic spectral estimators describe a generalising family of non-parametric spectral estimators (Percival & Walden, 1993; Walden et al., 1995; Walden, 2000). Define the complex valued demodulation process $\mathrm{Z}_t = \mathrm{X}_t \mathrm{e}^{i\omega t}$, which has corresponding auto-covariance function, $\gamma_{\mathrm{Z}}(\tau) = \gamma(\tau)\mathrm{e}^{i\omega\tau}$, and power spectral density $f_{\mathrm{Z}}(\omega') = f(\omega - \omega')$, in particular, $f_{\mathrm{Z}}(0) = f(\omega)$. Given an n-dimensional finite observation from $\{\mathrm{X}_t\}$, define X and Z as the column vectors $\mathrm{X} = (\mathrm{X}_0, \ldots, \mathrm{X}_{n-1})$ and $\mathrm{Z} = (\mathrm{Z}_0, \ldots, \mathrm{Z}_{n-1})$. In what follows, denote by superscript H the Hermitian transpose and by an overbar the complex conjugate.

DEFINITION 1. *Define a quadratic spectral estimator* $I_{\mathrm{Quad}}(\omega)$ *as*

$$I_{\mathrm{Quad}}(\omega) = \mathrm{Z}^{\mathrm{H}}\mathrm{Q}\mathrm{Z} = \sum_{s=0}^{n-1}\sum_{t=0}^{n-1} \bar{\mathrm{Z}}_s \mathrm{Q}_{s,t} \mathrm{Z}_t \tag{1}$$

where $\mathrm{Q}_{s,t}$ *is the* (s,t)*th element of the real valued and symmetric* $n \times n$ *matrix* Q.

Choosing the values of Q determines the properties of the estimator and, as we now demonstrate, encompasses the periodogram, lag-window, multitaper, and Welch estimators.

Example 1 *(Periodogram).* The periodogram, $I_n(\omega)$, with the chosen data-taper h_t such that $\sum_{t=0}^{n-1} h_t^2 = 1$ may be defined as the squared modulus of the discrete Fourier transform of X,

$$I_n(\omega) = \left|\sum_{t=0}^{n-1} h_t \mathrm{X}_t \mathrm{e}^{-i\omega t}\right|^2 = \sum_{s=0}^{n-1}\sum_{t=0}^{n-1} h_s h_t \mathrm{X}_s \mathrm{X}_t \mathrm{e}^{-i\omega(s-t)} = \sum_{s=0}^{n-1}\sum_{t=0}^{n-1} \bar{\mathrm{Z}}_s h_s h_t \mathrm{Z}_t = \mathrm{Z}^{\mathrm{H}}\mathrm{Q}\mathrm{Z}$$

The case of $h_t = n^{-1/2}$ corresponds to the choice of 'no taper' for the raw periodogram. The periodogram is therefore a quadratic spectral estimator with values $\mathrm{Q}_{s,t} = h_s h_t$ and Q is of unit rank.

Example 2 *(Lag-window).* Define a lag-window estimator as $I_{\mathrm{Lag}}(\omega) = \sum_{\tau=-n+1}^{n-1} (w_\tau \cdot \hat{\gamma})(\tau) e^{-i\omega\tau}$ where $\hat{\gamma}(\tau) = \sum_{t=0}^{n-|\tau|-1} h_{t+|\tau|}\mathrm{X}_{t+|\tau|} h_t \mathrm{X}_t$ and $\{w_\tau\}$ is the lag-sequence. We may re-write $I_{\mathrm{Lag}}(\omega)$ as

$$I_{\mathrm{Lag}}(\omega) = \sum_{s=0}^{n-1}\sum_{t=0}^{n-1} w_{(s-t)} h_s h_t \mathrm{X}_s \mathrm{X}_t \mathrm{e}^{-i\omega(s-t)}$$

so that $I_{\mathrm{Lag}}(\omega)$ is a quadratic spectral estimator where $\mathrm{Q}_{s,t} = w_{(s-t)} h_s h_t$ and Q may be up to full rank.

Example 3 *(Welch's estimator).* Welch's estimator is defined $I_{\mathrm{Welch}}(\omega) = M^{-1}\sum_{m=0}^{M-1} I_L^m(\omega)$ where $I_L^m(\omega)$ is the periodogram calculated from the mth length-L segment of the data. We may therefore define the $I_L^m(\omega)$ as periodograms calculated from the full length-n observation, so that

$$I_L^m(\omega) = \left|\sum_{t=0}^{n-1} h_t^m \mathrm{X}_t \mathrm{e}^{-i\omega t}\right|^2 \tag{2}$$

where, here, $\{h_t^m\}$ is a length L data taper that has been pre- and post-pended with mLp and $n - L(1 + mp)$ zeros, respectively, and p is the degree of segment overlap. From Example 1, we can write (2) in quadratic form $I_L^m(\omega) = \mathrm{Z}^{\mathrm{H}}\mathrm{Q}^m\mathrm{Z}$ where each Q^m is the outer product of $\{h_t^m\}$ with itself. Thus, $I_{\mathrm{Welch}}(\omega) = \mathrm{Z}^{\mathrm{H}}\mathrm{Q}\mathrm{Z}$ is a quadratic spectral estimator where $\mathrm{Q} = \frac{1}{M}\sum_{m=0}^{M-1}\mathrm{Q}^m$ and is of rank-M.

Example 4 *(Multitaper).* Multitaper estimators are defined as $I_{\mathrm{Multi}}(\omega) = \sum_{k=0}^{K-1} d_k I_n^k(\omega)$, where $\sum_{k=0}^{K-1} d_k = 1$ and the $I_n^k(\omega)$ are periodograms calculated with tapering sequence $\{h_{k,t}\}$, defined orthogonally over k so that $\sum_{t=0}^{n-1} h_{k,t}h_{k',t} = 0$ for $k \neq k'$. Similarly to Example 3, $I_n^k(\omega) = \mathrm{Z}^{\mathrm{H}}\mathrm{Q}^k\mathrm{Z}$ where each Q^k is the outer product of $\{h_{k,t}\}$ with itself, and $I_{\mathrm{Multi}}(\omega) = \mathrm{Z}^{\mathrm{H}}\mathrm{Q}\mathrm{Z}$ is a quadratic spectral estimator where $\mathrm{Q} = \sum_{k=0}^{K-1} d_k\mathrm{Q}^k$ and is of rank-K.

2. Statistical Properties

Assume Q has $\mathrm{tr}(\mathrm{Q}) = 1$, as in all previous examples, and is of rank K for $1 \leq K \leq n$. If Q is positive semi-definite then $I_{\mathrm{Quad}}(\omega)$ is guaranteed to be non-negative. It is common to assume positive semi-definite Q; although, as this excludes certain lag-window estimators (e.g. Politis & Romano, 1995, described in Example 6) we relax this assumption. The spectral decomposition of Q is $\mathrm{Q} = \mathrm{H}\mathrm{D}\mathrm{H}^{\mathrm{T}} = \sum_{k=0}^{K-1} d_k\mathrm{H}_k\mathrm{H}_k^{\mathrm{T}}$ where H is an orthogonal $n \times K$ matrix with columns $\mathrm{H}_0, \dots, \mathrm{H}_{K-1}$, and D is a $K \times K$ diagonal matrix with values $d_0, \dots, d_{K-1}$ such that $\sum_{k=0}^{K-1} d_k = 1$ because $\mathrm{tr}(\mathrm{Q}) = 1$. The $\mathrm{H}_0, \dots, \mathrm{H}_{K-1}$ and $d_0, \dots, d_{K-1}$ are the eigenvectors and eigenvalues of Q, respectively. By definition, the eigenvectors H_k respect unit square-summability and so are valid tapers. Thus, any quadratic spectral estimator (1) can be written as

$$I_{\mathrm{Quad}}(\omega) = \sum_{k=0}^{K-1} d_k \left| \sum_{t=0}^{n-1} h_{k,t}\mathrm{X}_t \mathrm{e}^{-i\omega t} \right|^2 = \sum_{k=0}^{K-1} d_k I_n^k(\omega)$$

which is identical to the form of the multitaper, but with the tapers defined by the eigenvectors of Q. Defining $\mathcal{H}_n^k(\omega) = \left|\sum_{t=0}^{n-1} h_{k,t}\mathrm{e}^{-i\omega t}\right|^2$, the expectation of each $I_n^k(\omega)$ is $\mathrm{E}[I_n^k(\omega)] = (f * \mathcal{H}_n^k)(\omega)$. Therefore,

$$\mathrm{E}[I_{\mathrm{Quad}}(\omega)] = \sum_{k=0}^{K-1} d_k \mathrm{E}[I_n^k(\omega)] = \left(f * \bar{\mathcal{H}}_n\right)(\omega) \tag{3}$$

where $\bar{\mathcal{H}}_n(\omega) = \sum_{k=0}^{K-1} d_k\mathcal{H}_n^k(\omega)$. For all $f(\omega)$ aside from white noise, $I_{\mathrm{Quad}}(\omega)$ is biased for finite n. As seen in Examples 1–4, $\bar{\mathcal{H}}_n(\omega)$ and thus

$$\mathrm{bias}[I_{\mathrm{Quad}}(\omega)] = \mathrm{E}[I_{\mathrm{Quad}}(\omega)] - f(\omega) = \left(f * \bar{\mathcal{H}}_n\right)(\omega) - f(\omega) \tag{4}$$

are dependent on the appointed tapers and lag-windows. The variance of $I_{\mathrm{Quad}}(\omega)$ is

$$\mathrm{var}[I_{\mathrm{Quad}}(\omega)] = \sum_{k=0}^{K-1} d_k^2 \mathrm{var}[I_n^k(\omega)] + 2\sum_{j<k} d_j d_k \mathrm{cov}[I_n^j(\omega), I_n^k(\omega)] = \sum_{k=0}^{K-1} d_k^2 \mathrm{var}[I_n^k(\omega)] \tag{5}$$

where the second equality results from $\mathrm{cov}[I_n^j(\omega), I_n^k(\omega)] = 0$ due to the orthogonality across the H_k. The covariance $\mathrm{cov}[I_n^k(\omega), I_n^k(\omega+\eta)]$, from which $\mathrm{var}[I_n^k(\omega)]$ in (5) is obtained, is calculated as

$$\mathrm{cov}[I_n^k(\omega), I_n^k(\omega+\eta)] = \left| \int_{-\pi}^{\pi} \bar{H}_k(\omega+\eta-\omega')H_k(\omega-\omega')f(\omega')\,\mathrm{d}\omega' \right|^2 + \underbrace{\left| \int_{-\pi}^{\pi} H_k(\omega+\eta+\omega')H_k(\omega-\omega')f(\omega')\,\mathrm{d}\omega' \right|^2}_{=\mathcal{O}(\log^2 n_0/n_0^2)} + \mathcal{O}(n_0^{-1}) \tag{6}$$

where $H_k(\omega) = \sum_{t=0}^{n-1} h_{k,t}\mathrm{e}^{-i\omega t}$ and n_0 is the number of non-zero elements in $\{h_{k,t}\}$. This is an extension of Theorem 5.2.5 of Brillinger (1975) to a general taper $\{h_{k,t}\}$ obtained by substituting Equation (211b) of Percival & Walden (2020). The $\mathcal{O}(\log^2 n_0/n_0^2)$ term is obtained in Krogstad (1982) for $h_{k,t} = n^{-1/2}$, and extends to the case of a general taper $\{h_{k,t}\}$ using Lemma 1 of Astfalck et al. (2024).

Assuming $\mathrm{var}[I_n^k(\omega)]$ is constant over k, $\mathrm{M}^{-1} = \sum_{k=0}^{K-1} d_k^2$ is an approximation of the variance reduction in $I_{\mathrm{Quad}}(\omega)$, as compared to the periodogram, and $I_{\mathrm{Quad}}(\omega)$ is an $\sqrt{\mathrm{M}}$-convergent estimator. When Q is positive semi-definite, $\mathrm{M}^{-1} \leq 1$. In actuality, $\mathrm{var}[I_n^k(\omega)]$ is not precisely constant over k, and so

M^{-1} is not the exact factor of variance reduction; however, it is useful in establishing certain properties in Remark 1, below. To arrive at a notion of bias-variance trade-off, we require a metric to describe the effective number of independent observations from the spectral estimator, that is, the effective sample size ζ. We seek ζ that is independent of ω and only depends on the parametrisation of the quadratic estimator and not the unknown $f(\omega)$. Similarly to how spectral bandwidth is commonly considered (e.g., see Percival & Walden, 2020), we define ζ for a white noise process with constant $f(\omega')$ as

$$\zeta = \frac{2\pi}{\int_{-\pi}^{\pi} \mathrm{cor}[I_z(\omega), I_z(\omega+\eta)] \, \mathrm{d}\eta} \tag{7}$$

where $I_z(\omega)$ is the spectral estimator $I_{\mathrm{Quad}}(\omega)$ of a white noise process. This definition of ζ is a re-expression of the common definitions in the stochastic process literature to the spectral domain, and as we will show, is simply the inverse of the effective bandwidth of a spectral estimator. Dividing (6) through by $f(\omega')$ and combining over the K components yields $\mathrm{cor}[I_z(\omega), I_z(\omega+\eta)]$ which is calculated

$$R(\eta) = \mathrm{M} \sum_{j,k=0}^{K-1} d_j d_k \left| \sum_{t=0}^{n-1} h_{j,t} h_{k,t} e^{-i\eta t} \right|^2 = \mathrm{cor}[I_z(\omega), I_z(\omega+\eta)] + \mathcal{O}(n_0^{-1}). \tag{8}$$

Substituting (8) into (7) yields $\zeta = 1/\mathcal{B}_{\mathrm{Quad}} + \mathcal{O}(n_0^{-1})$, where

$$\mathcal{B}_{\mathrm{Quad}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} R(\eta) \, \mathrm{d}\eta = \mathrm{M} \sum_{j,k=0}^{K-1} \sum_{t=0}^{n-1} d_j d_k h_{j,t}^2 h_{k,t}^2. \tag{9}$$

Equation (9) can be understood as the spectral bandwidth of $I_{\mathrm{Quad}}(\omega)$, i.e., the integral length scale of the spectral estimator, and is independent of ω. The effective sample size ζ admits values $1 \leq \zeta \leq n$.

PROPOSITION 1. *Given $\{\mathrm{X}_t\}$ satisfies Assumption 1, and $\omega \neq 0, \pm\pi$, the bias of a quadratic spectral estimator, $I_{\mathrm{Quad}}(\omega)$ is of order*

$$\mathrm{bias}[I_{\mathrm{Quad}}(\omega)] = \mathrm{E}[I_{\mathrm{Quad}}(\omega)] - f(\omega) = \mathcal{O}\left(\frac{\log \zeta}{\zeta}\right). \tag{10}$$

Proof. Re-write (4) as $\mathrm{bias}[I_{\mathrm{Quad}}(\omega)] = \sum_{k=0}^{K-1} d_k \{\mathrm{E}[I_n^k(\omega)] - f(\omega)\}$, and so $\mathcal{O}(\mathrm{bias}[I_{\mathrm{Quad}}(\omega)])$ is dominated by $\max_k[\{\mathcal{O}\{\mathrm{E}[I_n^k(\omega)] - f(\omega)\}]$. The maximum value for $\mathrm{E}[I_n^k(\omega)] - f(\omega)$ is achieved when the number of non-zero elements of $\{h_{k,t}\}$ are minimised. To maintain a length-n estimator, Equation (9) implies a minimum number of non-zero elements in each $\{h_{k,t}\}$ that scales with $\zeta \propto n/K$ as n grows. Following Lemma 1 of Astfalck et al. (2024), a taper with a minimum of ζ non-zero elements is bounded by $c_h \zeta^{-1/2}$ for some constant c_h, and so (10) is established following the result of Fejér (1910). □

Remark 1. To satisfy $\sum_{t=0}^{n-1} h_{k,t}^2 = 1$ with increasing n for each $\{h_{k,t}\}$, all tapers scale at maximum $h_{k,t} \propto \zeta^{-1/2}$. Further, to obtain the minimum of ζ non-zero elements in Proposition 1, the tapers are 'non-overlapping' so that $\sum_{t=0}^{n-1} h_{j,t}^2 h_{k,t}^2 = 0$ for $j \neq k$. As $\sum_{k=0}^{K-1} d_k = 1$, each $d_k \propto 1/K$ and consequently for a quadratic estimator, $I_{\mathrm{Quad}}(\omega)$, $\zeta\mathrm{M} \propto n$ as $\sum_{j,k=0}^{K-1} \sum_{t=0}^{n-1} d_j d_k h_{j,t}^2 h_{k,t}^2 = (\mathrm{M}\zeta)^{-1} \propto \sum_{k=0}^{K-1} \sum_{t=0}^{\zeta-1} (K\zeta)^{-2} = (K\zeta)^{-1} \propto n^{-1}$. By apportioning $\zeta = \mathcal{O}(n^{\alpha})$ and $\mathrm{M} = \mathcal{O}(n^{1-\alpha})$ the expected bias-variance trade-off is obtained. As $\mathrm{var}[I_{\mathrm{Quad}}(\omega)] = \mathcal{O}(\mathrm{M}^{-1})$, and from Proposition 1, the optimal value of α with respect to mean-squared-error is $\alpha \approx 1/3$, achieved by minimising $\mathrm{mse}[\cdot] = \mathrm{var}[\cdot] + \mathrm{bias}^2[\cdot]$ in terms of α for $n \to \infty$. Further regularity assumptions on $f(\omega)$, or parameterisations of the $h_{k,t}$, may reduce the bias in (10) and further lower the optimal value of α, thus allowing for more variance reduction. We show two instances of this in Examples 6 and 7.

Example 5 *(Parameterisation of Bartlett's estimator).* Bartlett's estimator arises as a special case of Welch's estimator and, as in Example 3, is equivalent to a quadratic estimator with $d_k = 1/K$ and $h_{k,t}$ is a length-L taper with values $L^{-1/2}$ pre- and post-pended with kL and $n - L(k+1)$ zeros. Substituting these values in (9) yields $\zeta = L$ and so $ML = n$ with equality, leading to the optimal choice of $\alpha \approx 1/3$.

Example 6 *(Parameterisation of an infinite-order flat-top lag-sequence).* Politis & Romano (1995, 1999) specify the infinite-order flat-top lag-sequence for univariate $\{\mathrm{X}_t\}$, with

$$w(\tau) = \begin{cases} 1 & |\tau| \leq 1/a \\ g(\tau; a) & 1/a < |\tau| \leq c/a \\ 0 & |\tau| > c/a \end{cases}$$

where $g(\tau; a)$ is a continuous real-valued function, defined over τ and parameterised by a, that satisfies $g(\tau; a) = g(-\tau; a)$, $g(\tau; 1) = g(\tau/a; a)$, $g(1/a; a) = 1$ and $g(c/a; a) = 0$. If $f(\omega)$ has r bounded and continuous derivatives, then $\mathrm{bias}[I_{\mathrm{Quad}}(\omega)] = \mathcal{O}(\zeta^{-r})$. Calculated similar to Remark 1 it is asymptotically optimal to choose $\alpha = 1/(2r+1)$, where Assumption 1 states $r \geq 1$.

Example 7 *(Parameterisation of a sinusoidal multitaper).* Sinusoidal multitapers are discrete approximations to minimum bias multitapers, specified with weights $d_k = 1/K$ for $k \in \{0, \ldots, K-1\}$ by

$$h_{k,t} = \left(\frac{2}{n+1}\right)^{1/2} \sin\left\{\frac{(k+1)\pi t}{n+1}\right\},$$

see Riedel & Sidorenko (1995). This leads to $\mathcal{B}[I_{\mathrm{Quad}}(w)] = (K+1)/(n+1)$ (Walden et al., 1995), thus $\zeta \propto n/K$, $\mathrm{M} = K$ and $\zeta\mathrm{M} \propto n$ as desired. As shown in Riedel & Sidorenko (1995), the bias of a Sinusoidal multitaper is $\mathcal{O}(\zeta^{-2})$, and therefore suggests an optimal choice of $\alpha = 1/5$.

Remark 2. Studying the theory of quadratic estimators provides a principled framework with which to combine classes of non-parametric spectral estimators that are typically siloed. For instance, multitaper estimators based on Slepian sequences with large K are known to re-introduce bias as K increases; but without increasing K we cannot resolve variance. Often in scientific application there are requirements for variance to be resolved under some threshold and may require a choice of α that does not necessarily adhere to asymptotic arguments. For example, Welch and multitaper methods can be combined by segmenting the time-series, calculating multitaper estimates on each segment for some fixed K, and then averaging the segments to resolve variance; see Schmidt (2022) for an example. Such an estimator remains a quadratic estimator, and so we may obtain the statistical properties as described herein. Historically, quadratic estimators were predominantly discussed to motivate the use of constructed multitaper estimators whereby the $\{h_{k,t}\}$ are specified directly rather than as a consequence of specifying Q.

3. Bias correction of quadratic spectral estimators

3.1. Bias correction

Assume a family of bases, such that the power spectral density may be represented as $f(\omega; \vartheta) = \sum_{s=0}^{S-1} a_s b_s(\omega)$ where $b_s(\omega)$ is the sth basis function with corresponding auto-correlation function $\rho_s(\tau)$ obtained via an inverse Fourier transform, and $\vartheta = (a_0, \ldots, a_{S-1})$. Astfalck et al. (2024) proposed a methodology to debias Welch's estimator to $f(\omega)$ whereby the bases are intentionally biased, basis coefficients are inferred by fitting the biased bases to the estimate of $f(\omega)$ to obtain the $\hat{a}_s$ and a bias corrected estimate of $f(\omega)$ is obtained from $\sum_{s=0}^{S-1} \hat{a}_s b_s(\omega)$. We now present the theory to extend these ideas to the general class of quadratic spectral estimators, subsuming the methodology of Astfalck et al. (2024).

Define the Fourier frequencies $(\omega_0, \ldots, \omega_{n-1}) = 2\pi n^{-1}(-\lfloor n/2 \rfloor, \ldots, \lceil n/2 \rceil - 1)$, the vector $\mathrm{b}_s = \{b_s(\omega_0), \ldots, b_s(\omega_{n-1})\}$ containing the evaluations of the sth basis over the Fourier frequencies, and the $S \times n$ matrix $\mathrm{B} = (\mathrm{b}_0^{\mathrm{T}}, \ldots, \mathrm{b}_{S-1}^{\mathrm{T}})$ that stores these evaluations over all S bases. We bias the bases so that, $\breve{b}_s(\omega) = (b_s * \bar{\mathcal{H}}_n)(\omega)$, where $\bar{\mathcal{H}}_n(\omega)$ defines the bias in $I_{\mathrm{Quad}}(\omega)$, see (3). We calculate $\breve{\mathrm{b}}_s = \{\breve{b}_s(\omega_0), \ldots, \breve{b}_s(\omega_{n-1})\}$ via

$$\breve{b}_s(\omega) = (b_s * \bar{\mathcal{H}}_n)(\omega) = 2 \times \mathrm{Re}\left[\sum_{k=0}^{K-1} d_k \left\{\sum_{\tau=0}^{n-1} \left(\sum_{t=0}^{n-\tau-1} h_{k,t} h_{k,t+\tau}\right) \rho_r(\tau) e^{-i\omega\tau}\right\}\right] - 1 \tag{11}$$

and further define the $S \times n$ matrix of biased bases evaluations as $\breve{\mathrm{B}} = (\breve{\mathrm{b}}_0^{\mathrm{T}}, \ldots, \breve{\mathrm{b}}_{S-1}^{\mathrm{T}})$.

We model $I_{\text{Quad}}(\omega) = \sum_{s=0}^{S-1} a_s \check{b}_s(\omega) + \epsilon(\omega)$, where $\text{var}[\epsilon(\omega)] = \text{var}[I_{\text{Quad}}(\omega)]$ and for sufficiently large M, the central limit theorem implies that $\epsilon(\omega)$ may be considered Gaussian. We expect there to be non-ignorable correlations between the frequencies for certain choices of Q, which is well represented by $R(\eta)$ in (8). Different to Astfalck et al. (2024), we specify a circulant correlation matrix W with (i, j)th values $\mathrm{W}_{i,j} = \text{cor}[I_{\text{Quad}}(\omega_i), I_{\text{Quad}}(\omega_j)] \approx R(\omega_i - \omega_j)$ and solve the weighted least squares problem

$$\hat{\vartheta} = \arg\min_{\vartheta} \left\{ (\mathrm{I}_{\text{Quad}} - \check{\mathrm{B}}^{\mathrm{T}}\vartheta)^{\mathrm{T}} \mathrm{V}^{-1} (\mathrm{I}_{\text{Quad}} - \check{\mathrm{B}}^{\mathrm{T}}\vartheta) \right\} \tag{12}$$

where $\mathrm{I}_{\text{Quad}} = \{I_{\text{Quad}}(\omega_0), \ldots, I_{\text{Quad}}(\omega_{n-1})\}$, $\mathrm{V} = \Gamma \mathrm{W} \Gamma$, and Γ is a diagonal matrix with (i,i)th values $\Gamma_{i,i} = \text{sd}[I_{\text{Quad}}(\omega_i)]$ approximated by $I_{\text{Quad}}(\omega_i)$. This follows from Theorem 2 of Astfalck et al. (2024) which may be generalised to this case by bounding the tapers of the quadratic estimator as per Lemma 1 of Astfalck et al. (2024). The basis coefficients $\hat{\vartheta}$ in (12) are analytically available via the standard weighted least squares solution, and can be constrained to yield a non-negative estimate of the spectral density via non-negative least squares solutions. The validity of (12) as a solution is predicated on the appropriate assumption of Gaussian $\epsilon(\omega)$, implying a sufficiently large M. For small M, these assumptions break, and worse performance in terms of mean-square-error can be observed. In practice, we find $M \approx 16$ to be the minimum advisable value. Especially for short time-series, this choice should be considered as large M implies a reduction in ζ and so good resolution of the spectral density may be difficult to obtain.

Remark 3. Our methodology is predicated on the availability of $\bar{\mathcal{H}}_n(\omega)$, simply calculated via (3) for all quadratic estimators. The ability to extend this methodology to other estimators is similarly predicated on the ability to calculate $\bar{\mathcal{H}}_n(\omega)$. For instance, McElroy & Politis (2024) propose an estimator to the spectral density at frequencies near-zero with a better convergence rate, in effect by altering the lag-window to be symmetric about zero at such frequencies. Therein, $\bar{\mathcal{H}}_n(\omega)$ is obtainable by considering the family of lag-windows and similarly our methodology can be extended to such estimators.

3.2. *Computation*

The computational order of the bias correction of the quadratic estimator is $\mathcal{O}\{\max(nS^2, nKS\log n)\}$, where S is the number of bases and K is the rank of Q. The first term is the computation required to calculate the solution to (12) and the second term is the computation required to calculate the $\check{b}_s(\omega)$ in (11). For real-valued data $f(\omega)$ is symmetric, to ensure numerical stability, S should scale at maximum with half the effective sample size ζ (in practice we find $S = \zeta/4$ works well) so that $S = \mathcal{O}(n^\alpha)$, where α has a maximum value of $\alpha = 1/3$ (see Remark 1) and so $\mathcal{O}(nS^2) = \mathcal{O}(n^{5/3})$. The matrix V is circulant and so the inverse in (12) is calculated with order $\mathcal{O}(n \log n)$ and does not affect the computational order of (12). For the computation in (11), if $K \propto M = \mathcal{O}(n^{2/3})$, as in Welch's estimator and certain multitapers, then $\mathcal{O}(nKS\log n) = \mathcal{O}(n^2 \log n)$; if $K = n$ as in lag-windows, $\mathcal{O}(nKS\log n) = \mathcal{O}(n^{7/3}\log n)$. At first, these rates may seem large, but they are a result of generalising the theory so as to establish the statistical properties. Instead, (11) may be computed faster for Welch and lag-window estimates: Welch estimates follow Astfalck et al. (2024) and have order $\mathcal{O}(n\log n)$; lag-windows are computed similar to Equation (9) of Sykulski et al. (2019) with the addition of a lag-sequence and have order $\mathcal{O}(n^{5/3}\log n)$. As opposed to Astfalck et al. (2024) who retain $\mathcal{O}(n\log n)$, we find a worse-case scenario rate of $\mathcal{O}(n^2\log n)$, corresponding to a multitaper bias correction. Although this is larger than $\mathcal{O}(n\log n)$, computation of an $\mathcal{O}(n^2\log n)$ estimator is still possible for $n = \mathcal{O}(10^5)$ on a modern laptop computer.

3.3. *Flexible basis functions*

In Astfalck et al. (2024) the $b_s(\omega)$ were defined as rectangular bases, in effect representing a Riemannian approximation to $f(\omega)$. Assumptions on the order of differentiability of $f(\omega)$ can inform the bias that results from the Riemannian approximation: $\mathcal{O}(S^{-1})$ if $f(\omega)$ is of bounded first derivative, and $\mathcal{O}(S^{-2})$ if $f(\omega)$ is of bounded variation. The standard Riemannian basis is a polynomial of order 0 and the integrated mean squared errors can be improved upon by employing basis functions with increasing polynomial order and thus increasing degrees of differentiability. For instance, the trapezoidal rule has integrated mean square error $\mathcal{O}(S^{-3})$ and corresponds to a polynomial basis of order 1, and Simpson's rule has integrated mean square error $\mathcal{O}(S^{-4})$ and corresponds to a polynomial basis of order 2.

For the sth basis, $b_s(\omega)$, define ω_s^{c} as the basis centre, δ_s as the basis width, and assume $b_s(\omega)$ to be p-times differentiable. We define $\rho_s(\tau)$, the auto-correlation function that corresponds to $b_s(\omega)$, as $\rho_s(\tau;\omega_s^{\mathrm{c}},\delta_s) = \{\delta_s^{-\frac{1}{p+1}}\operatorname{sinc}(\delta_s\tau)\exp(i\omega_s^{\mathrm{c}}\tau)\}^{p+1}$ where $\operatorname{sinc}(\cdot)$ is the normalised sinc function. For $p = 0$, basis $b_s(\omega)$ is $b_s^0(\omega) = \operatorname{rect}(\omega/\delta_c - \omega_c)$, where $\operatorname{rect}(\cdot)$ is the rectangular function. Higher orders are obtained via the recursion $b_s^p(\omega) = (b_s^{p-1} * b_s^0)(\omega)$ where the $b_0^p(\omega),\ldots,b_S^p(\omega)$ define a family of p-times differentiable functions, analogous to p-order B-splines defined over a circular domain. Choosing large p imposes an assumption on the regularity conditions of $f(\omega)$; if p is chosen correctly then an optimal rate of convergence may be guaranteed. However, if p is overestimated then this may impose an assumption of smoothness on the underlying function that is inappropriate and so may re-introduce bias. In practice when $f(\omega)$ is unknown, p must be estimated; see Politis & Romano (1995), who demonstrate an empirical methodology to estimate p from the empirical auto-covariance sequence $\hat{\gamma}(\tau)$. If we define $\delta_s = \delta$ for all s, our estimator calculates estimates of the spectral density over an even partition in frequency; whilst intuitively similar to quadratic estimators this is not a required assumption. Astfalck et al. (2024) provide an example with unevenly spaced bases with $p = 0$, and we may extend this idea for $p > 0$.

4. Simulation study with canonical AR(4) model

We demonstrate our general bias correction technique on lag-window, multitaper and Welch estimates of the canonical AR(4) model of Percival & Walden (1993), defined as $\mathrm{X}_t = \sum_{p=1}^{4}\phi_p\mathrm{X}_{t-p} + \epsilon_t$ where $\epsilon_t \sim \mathcal{N}(0,\sigma^2)$ and $\{\phi_1,\phi_2,\phi_3,\phi_4,\sigma\} = \{2.7607, -3.8106, 2.6535, -0.9238, 1\}$. We first present a bias correction to a single sample of length $n = 2^{14}$ for each of the three estimators, shown in Figure 1. We set $p = 1$ for the bases, and parameterise each of the three estimators so as to give similar expected bias: the lag-window estimator uses a modified Daniell of width $M = 2^5$; the multitaper estimator specifies the tapers as $K = 2^5$ Slepian sequences with a time-bandwidth parameter of 2^4; and the Welch estimator sets $L = 2^9$, $M = 2^5$, with zero-overlap and employs a Hamming data-taper on each segment. The number of basis are set as $S = n/4\mathrm{M} = 2^7$, approximately $\zeta/4$ for each estimator; in effect, this choice of S downscales the effective sample size of the Fourier frequencies by a factor of 2. Estimates are plotted in Figure 1 by the grey solid lines with their respective expectations, calculated as per (3), plotted with the grey dashed lines. The bias corrected estimates are all given by the black solid lines, and the true $f(\omega)$ of the AR(4) model is given by the black dashed line. The computational time to bias correct the lag-window and multitaper estimates was ~ 200 ms, and ~ 5 ms for the Welch estimate. In all cases, the debiased estimator provides a better estimate of the true spectrum across all frequencies. Finally, in Figure 2 we show performance of the bias corrected estimators over a 1000-member ensemble of simulations and as

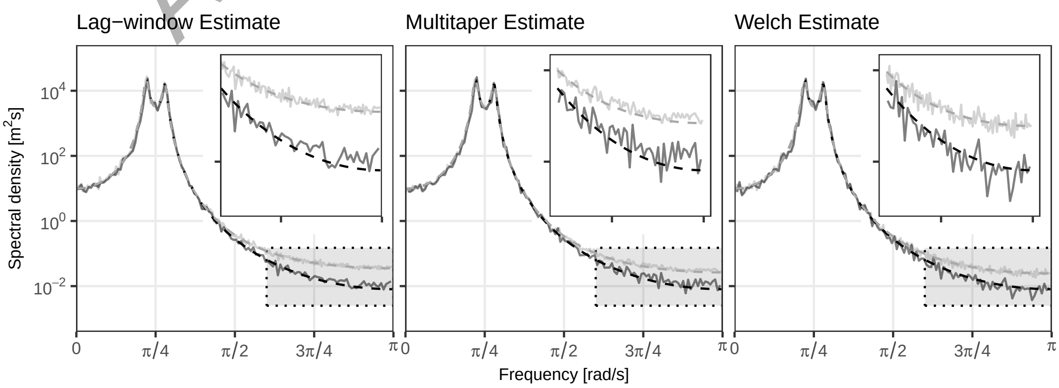

Fig. 1: Standard and debiased lag-window, multitaper and Welch estimates of a random AR(4) process. The standard and debiased estimates are shown by the solid grey and black lines, respectively; the true process spectrum is shown by the black dashed line and the expectation of the standard estimator by the grey dashed line. Plot insets correspond to the shaded dotted boxes. Parameterisations of the AR(4) process and the estimators are given in the main text.

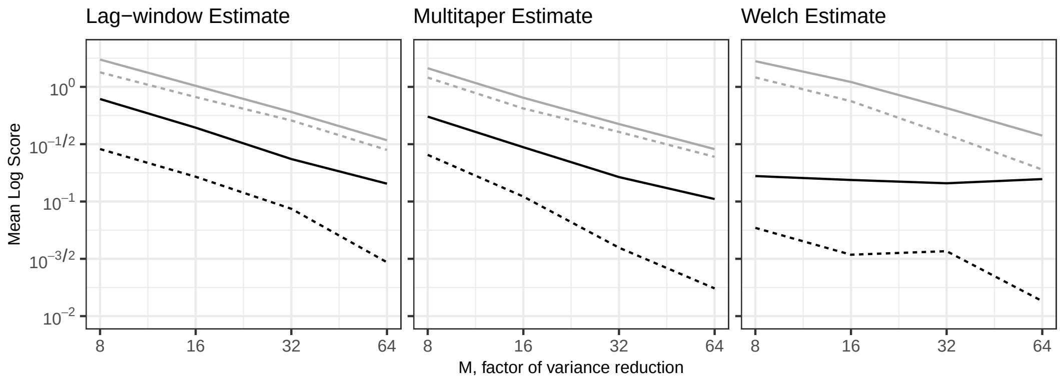

Fig. 2: Estimates of bias (black) and root-mean-squared-error (grey) of the standard (solid) and debiased (dashed) lag-window, multitaper and Welch estimates as a function of the factor of variance reduction, M. Each graphed point is obtained by calculating the metric from a 1000 member ensemble, for each frequency, and aggregating over frequency by the mean log value. Parameterisations of the AR(4) process and the estimators is given in the main text.

a function of M, the factor of variance reduction, with fixed $n/\mathrm{M} = 2^9$ and $S = n/4\mathrm{M} = 2^7$ for each simulation. An `R` implementation of the code is available at `github.com/astfalckl/dquad`.

ACKNOWLEDGEMENT

All authors are supported by the ARC ITRH for TIDE, Grant No. IH200100009.

REFERENCES

ASTFALCK, L. C., SYKULSKI, A. M. & CRIPPS, E. J. (2024). Debiasing Welch's method of spectral density estimation. *Biometrika* , asae033.

BRILLINGER, D. R. (1975). *Time series: data analysis and theory*. Holt, Rinehart and Winston Inc.

FEJÉR, L. (1910). Lebesguessche Konstanten und divergente Fourierreihen. *Journal für die Reine und Angewandte Mathematik* **138**, 22–53.

KROGSTAD, H. E. (1982). On the covariance of the periodogram. *Journal of Time Series Analysis* **3**, 195–207.

MCELROY, T. & POLITIS, D. N. (2024). Estimating the spectral density at frequencies near zero. *Journal of the American Statistical Association* **119**, 612–624.

PERCIVAL, D. B. & WALDEN, A. T. (1993). *Spectral analysis for physical applications*. Cambridge University Press.

PERCIVAL, D. B. & WALDEN, A. T. (2020). *Spectral analysis for univariate time series*. Cambridge University Press.

POLITIS, D. N. & MCELROY, T. S. (2019). *Time series: A first course with bootstrap starter*. Chapman and Hall/CRC.

POLITIS, D. N. & ROMANO, J. P. (1995). Bias-corrected nonparametric spectral estimation. *Journal of time series analysis* **16**, 67–103.

POLITIS, D. N. & ROMANO, J. P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *Journal of Multivariate Analysis* **68**, 1–25.

RIEDEL, K. S. & SIDORENKO, A. (1995). Minimum bias multiple taper spectral estimation. *IEEE Transactions on Signal Processing* **43**, 188–195.

SCHMIDT, O. T. (2022). Spectral proper orthogonal decomposition using multitaper estimates. *Theoretical and Computational Fluid Dynamics* **36**, 741–754.

SYKULSKI, A. M., OLHEDE, S. C., GUILLAUMIN, A. P., LILLY, J. M. & EARLY, J. J. (2019). The debiased Whittle likelihood. *Biometrika* **106**, 251–266.

WALDEN, A., MCCOY, E. & PERCIVAL, D. (1995). The effective bandwidth of a multitaper spectral estimator. *Biometrika* **82**, 201–214.

WALDEN, A. T. (2000). A unified view of multitaper multivariate spectral estimation. *Biometrika* **87**, 767–788.

[*Received on 22 October 2024*]