On Digital Signal Processing of Time Series for Spectrum Estimation

Dazhen Gu, Jacob Rezac, Xifeng Lu and Daniel Kuester

***Abstract*—We present a study of power spectral density (PSD) estimation from data sampled in the time domain. This work was motivated by our recent development of digital radiometry, where radiation spectra were obtained by processing the digitally sampled signal. The PSD estimation can be generalized by a quadratic estimator and minimization of mean squared error of the estimator leads to the optimal window choice. The bounds of the variance and the bias are formulated in order to quantify the uncertainty associated with non-ideal PSD estimation in digital signal processing. Windowed estimates of spectrum measurements are presented for comparison in terms of computational efficiency and amplitude measurement precision. A few examples on real and simulated data are shown for comparison.**

***Index Terms*—Measurement, power spectral density, quadratic estimation, spectrum analysis, time series, uncertainty.**

I. Introduction

Spectrum estimation deals with transforming a finite-length data series in time domain to determine the distribution of signal amplitude and phase in frequency domain. The data series can originate from different domains. For example, data in time domain periodically sampled by a digitizer or data in spatial domain recorded by an array of receivers. Spectrum estimation essentially allows representation of the signal in time domain (t) by the signal in frequency domain (f) or representation of the signal in spatial domain (x) by the signal in spatial-frequency domain (k). Multi-dimensional signals are also of use and can be analyzed with similar techniques. Spectrum estimation holds significance in many applications, including speech recognition [1], [2], biomedical diagnosis [3], geophysics [4], [5], radio astronomy [6], [7], radio-frequency spectrum management [8], source detection by radars and sonars [9], etc.

Regardless of the notation difference, either t and f or x and k form a Fourier transform pair. They can be treated in the same way. Without losing generality, we limit our discussion to one-dimensional signals in the time-frequency domain. Extensions to higher dimensions and different domains can be developed by following similar principles. In practice, we can only obtain data series with a finite length. As a result, the determination of the signal over frequency is, in general, an approximation. An exception is when the signal only contains a limited number of sinusoidal tones with its highest frequency below the Nyquist sampling frequency. If these tone frequencies are known *a priori*, proper sampling schemes can be designed for spectrum estimation to perfectly reconstruct the signal in the frequency domain. This exception is an ideal case. In reality, we often handle a set of unknown tones in conjunction with noise signals or band-limited noise and noise-like signals[1].

Spectrum estimation can generate multiple products; namely amplitude, tone frequency and phase. Depending on the nature of the signal, all or a part of the products are of interest. For example, all three parameters fully characterize a multi-tone signal and only the amplitude and the frequency are needed for a band-limited wide-sense stationary (WSS) signal. On one hand, identification of frequencies is critical in many applications, particularly true for signals consisting of multiple tones. Uncertainty for the frequency evaluation has been investigated in the past. Various spectrum estimation methods were compared in terms of the measurement uncertainties of frequencies [10]. On the other hand, accurate determination of the amplitude is also important. Applications can be found in radiation monitoring, channel occupation detection for opportunistic communication, and digital radiometry.

Published literature mostly deals with the bias and the variance of the amplitude of the estimated spectra. The basic periodogram and correlogram methods rely on the fundamental discrete Fourier transform (DFT) [11]. Spectrum estimation by these methods are plagued by the spectral leakage associated with large side lobes. Refinements have been developed by applying optimized window functions to suppress the side lobe effect.

II. Context

A spectrum analyzer is one of the most common instruments for high-frequency spectrum measurements. These consist of an analog frequency conversion stage, which sets and sweeps the center frequency over time, with filtering to present a band-limited signal for further measurements. Historically, only analog power measurements were performed at each center frequency. Modern instruments, however, which have become known as signal analyzers, have transitioned towards increasingly digital detection and analysis, allowing improved flexibility, speed, and frequency resolution. In our own past work, we developed a digital radiometer for noise metrology applications, with an architecture similar to modern signal analyzers [12], [13]. With a digital backend, the radiometer

D. Gu, X. Lu and D. Kuester are with the Shared Spectrum Metrology Group, National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: dazhen.gu@nist.gov).

J. Rezac is with the High-Speed Waveform Metrology Group, National Institute of Standards and Technology, Boulder, CO 80305 USA

Manuscript received XXXX XX, 2023.

[1]Here noise signals fall into the category of a wide-sense stationary (WSS) signal, for which the first and the second moments are deterministic.

achieved sideband separation, higher detection speed and finer spectral resolution. Improvements are about several orders of magnitude better in comparison to the analog counterpart. Nevertheless, the digital radiometer is fundamentally different from the analog radiometer in the frequency-dependent power detection of WSS signals. The analog radiometer provides a direct frequency-domain power measurement, while the digital radiometer relies on the digital signal processing (DSP) of the time series. As a consequence, it is important to study the measurement uncertainty introduced by the spectrum estimation with DSP.

Measurement uncertainty holds significance in metrology and other applications. Any measurement process is governed by an underlying physical model. Such a physical model defines parameters that characterize the measurement apparatus and their relations to the end parameter derived from measurements. The purpose of the calibration procedures is to determine these apparatus parameters. The calibration is practically implemented by use of a set of reference standards. The physical characteristics of these reference standards are only known to a limited extent. Measurement uncertainty quantifies such a limitation. The uncertainty of apparatus parameters ultimately determines the uncertainty of the end parameter. In spectrum measurements, the end parameter is spectral power or power spectral density (PSD).

The uncertainty associated with the NIST analog radiometer amounts to about 1% of the nominal value of the PSD level. The digital radiometer shares very similar frontend components with the analog radiometer. The differences lies in the replacement of the analog detection backend with a digital sampler. As mentioned above, DSP is conducted on the digitally sampled data in the time domain to obtain spectral power in the frequency domain. Even for DSP with optimized window functions, the spectrum calculation is an approximation. DSP optimization often trades closer approximation with more computational expense. We need to quantitatively evaluate such an approximation so that it can be compared to the uncertainty associated with other factors. To that end, an educated decision can be made to choose proper DSP techniques tailored for specific applications.

Returning back to the uncertainty and calibration issues, spectrum measurements by digitally processed time series entail data acquisition from an instrument and a subsequent DSP step. The DSP step can be considered as a virtual instrument that accepts data inputs in time domain and produces data outputs in frequency domain. A typical calibration can correct for the systematic errors associated with the instrument and the uncertainty pertaining to the systematic error coefficients is also taken into account [14]. This instrument calibration process is documented in great detail in our previous digital radiometry work [13], which we won't repeat in this study. The calibration, however, cannot fully account for the residual bias in the DSP step. It will become evident that the residual bias isn't invariant and depends not only on the DSP method but also on the measured signals, whether arising from a calibration reference or from an unknown signal source. In this paper, we attempt to develop an analytic method to quantify such a residual bias. This topic will be revisited in Sections III and IV.

As the discretely sampled digital detection becomes mainstream for scientific instruments, a general user often has a choice to select a window function for processing the sampled signal [15]. Subsequently, DSP is implemented in the instrument by use of the user-defined window to generate the end results of measurements. These features are fully integrated in modern signal analyzers. However, some fundamental questions remain unclear; 1) which type of window function is needed to reproduce the true signal, 2) will the DSP step introduce significant measurement uncertainty, and 3) how do we make tradeoffs between measurement speed and accuracy. We attempt to resolve these questions in this paper. In what follows, we present a rigorous analysis of PSD estimates in Section III in order to address Question 1). Calculation of PSD bias introduced by DSP is presented in Sections III and IV so that the pertinent uncertainty can be quantified to answer Question 2). In Section V, we provide three examples to evaluate the practical impact of different window functions, shedding some light on Question 3).

III. Estimation of Power Spectral Densities

Loosely speaking, the PSD of a time series is the Fourier transform of its autocorrelation. Indeed, if x_n, $n \in \mathbb{Z}$ is a stationary, ergodic, mean-zero time series with autocorrelation

$$r_n = \left\langle x_{n'} x^*_{n'-n} \right\rangle, \quad n', n \in \mathbb{Z}, \tag{1}$$

then its power spectral density is

$$S(f) = \sum_{n=-\infty}^{\infty} r_n \exp\left(-2\pi i f n\right), \quad f \in \mathbb{R}. \tag{2}$$

Notations $\mathbb{Z}$ and $\mathbb{R}$ indicate integers and real numbers, respectively. The operator $\langle\cdot\rangle$ means the expected value or equivalently the time average for an ergodic process. Power spectral densities do not always exist as standard functions (though they may exist as generalized functions similar to the Dirac delta distribution), but in this article we assume that $S(f)$ does (see [16] for a more rigorous discussion of this).

In practice, we are dealing with a time series of a finite length. Here, we consider the sampling time t_s is normalized to 1 and the frequency range of the signal of interest is within $\left[-\frac{1}{2}, \frac{1}{2}\right]$. The PSD estimation problem is to estimate $S(f)$, for frequencies $f \in \left[-\frac{1}{2}, \frac{1}{2}\right]$, from some $N \times 1$ vector of measurements $\mathbf{x}$ whose nth element is x_n.

One of the most common estimators of the PSD [11], known as the periodogram, is defined by

$$\hat{S}_{\mathrm{p}}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x_n \exp\left(-2\pi i f n\right) \right|^2. \tag{3}$$

This is an intuitive estimator and it can be rapidly computed at specific frequencies with a fast Fourier transform. However, it is only asymptotically unbiased, meaning that N must be very large in order for $\left\langle \hat{S}_{\mathrm{p}}(f) \right\rangle \approx S(f)$. Moreover, the variance of this estimator is proportional to $S^2(f)$ [11], even for large values of N!

In addition, the PSD can also be calculated by the previously mentioned definition from the estimated autocorrelation function.

$$\hat{S}_{\mathrm{c}}(f) = \sum_{n=-(N-1)}^{N-1} \hat{r}_n \exp\left(-2\pi i f n\right). \tag{4}$$

This is known as the correlogram estimate. Here the estimated autocorrelation $\hat{r}_n$ is computed from the time series with a finite length by

$$\hat{r}_n = \frac{1}{N} \sum_{n'=n+1}^{N-n} x_{n'} x^*_{n'-n}. \tag{5}$$

It can be shown that the periodogram and the correlogram are equivalent [11]. As such, the correlogram estimate suffers similar limitations.

One way to address these limitations is through a tapered periodogram,

$$\hat{S}_{\mathrm{w}}(f) = \left| \sum_{n=0}^{N-1} x_n w_n \exp\left(-2\pi i f n\right) \right|^2, \tag{6}$$

where w_n, the taper function, fulfills the normalization condition $\|\mathbf{w}\|^2 = 1$. The taper weights the estimate such that the bias of $\hat{S}_{\mathrm{w}}$ is lower than the bias of $\hat{S}_{\mathrm{p}}$ with the same number of samples. Such tapered estimates are typically useful, but choosing a good taper is difficult when measuring a time series with unknown statistical qualities. The following sections outline the process of choosing a good taper.

A. Spectrum Function Interpolation

In this subsection, we describe PSD estimates that are defined as functions on the entire band of interest $B = [-\frac{1}{2}, \frac{1}{2}]$ with pre-specified analytical properties. These estimates allow us to incorporate *a priori* information about the true PSD (*e.g.*, the existence of derivatives to a certain order) in a systematic manner. We will demonstrate that this framework can be studied with some of the same tools which have been used in previously described PSD estimation theory. The framework also suggests that new tools may be beneficial when strong *a priori* information exists.

There has been significant previous effort in deriving tapers for PSD estimation. We deviate from the typical discussion by treating a PSD estimate $\hat{S}(f)$ as an interpolating function of $S(f)$ which meets certain properties. When deriving PSD estimates, authors often assume that $S(f)$ varies slowly over subsets of B. We interpret this as an assumption that the true $S(f)$ can be well-estimated by piecewise constant functions. Such an assumption is often reasonable, but there are many cases in which it can be improved; if there are strong reasons to suspect that the true $S(f)$ has high-order derivatives which exist on all of B, then it is better to estimate it with a function which shares that property. Here we start with a general polynomial interpolation without specific assumptions. This approach in turn enables us to develop a complete model for mean-squared error (MSE) that incorporates a key component that has been summarily dismissed by prior work. The formulations of MSE are detailed in the following subsection.

In this work, we build PSD estimates with pre-specified continuity properties by first breaking B into K non-overlapping subsets $B_k \subset B$ over which $\hat{S}(f)$ is defined with piecewise-polynomial interpolants. We continue by taking $\hat{S}(f)$ to be a function of the form,

$$\hat{S}(f) = \sum_{k=0}^{K-1} \sum_{p=0}^{P} c_k^{(p)} \varphi_k^{(p)}(f), \tag{7}$$

where P is the highest order of polynomial interpolants. $\varphi_k^{(p)}$ are piecewise polynomials which are only non-zero on non-overlapping subsets $B_k \subset B$ and $c_k^{(p)}$ are constants to be estimated. For example, if $[-1/2, 1/2]$ is uniformly discretized on a grid with spacing ΔF with the points $f_k = -1/2 + (k + 1/2)\Delta F$ for $k = 0, \ldots, (1/\Delta F - 1)$, then we might chose $B_k = [f_k - \Delta F/2, f_k + \Delta F/2]$ and use piecewise constants with support only on each B_k; here, $P = 0$. We focus primarily on this example throughout the article, because there are closed-form bounds helpful in its analysis, but continue our derivation in more generality throughout this and next subsections. We believe new tools might be developed for the general problem in the future.

Piecewise polynomial interpolants of this type are studied commonly in the context of the finite element method [17]. Therein, piecewise Lagrange polynomials are often used to locally approximate the solution of differential equations. These polynomials are constructed to have pre-specified continuity properties across grid discretizations. For example, piecewise linear functions, and higher interpolants, constructed in this manner are designed to be continuous across all of $B = [-1/2, 1/2]$. To ensure this, they the function evaluated at the rightmost grid point of a subinterval B_{k-1} equals the function evaluated at the leftmost grid point of a subinterval B_k. Referring to (7) and the grid described in the example from the above paragraph, this results in the constraint $\sum_{p=0}^{P} c_{k-1}^{(p)} \varphi_{k-1}^{(p)}(f_{k-1} + \Delta F/2) = \sum_{p=0}^{P} c_k^{(p)} \varphi_k^{(p)}(f_k - \Delta F/2)$ for all $k = 1, \ldots, K-1$. More generally, the form of interpolant described in (7) consists of $P > 1$ polynomials of degree P on each subset and $\varphi_k^{(p)}$ are chosen so that $\frac{d^{p-1}\hat{S}(f)}{df}$ exists everywhere on B. We refer to [17] for more details about these piecewise polynomial estimates, as used in the finite element method.

B. Mean Squared Error Estimates

One benefit of explicitly describing PSD estimates with interpolating functions is that we can estimate the expected error of a PSD estimate across the entirety of B. We re-emphasize that this differs from the standard discussion, which ignores the approximation error through the interpolating functions described above. We represent the difference between our estimate (6) and the true PSD can be upper bounded by the difference between (6) and the interpolation of the true PSD PLUS the difference between the true PSD and the interpolation of the true PSD.. Using Lagrange interpolation, we can write

$$(\mathbb{P}^{(p)} S)(f) = \sum_{k=0}^{K-1} \sum_{p=0}^{P} S(f_k^{(p)}) \varphi_k^{(p)}(f), \tag{8}$$

where $f_k^{(p)}$ are appropriate frequency points for the interpolation order p. We can bound this expected MSE between S and $\hat{S}^{(p)}$ by

$$\int_B \left(S(f) - (\mathbb{P}^{(p)}S)(f)\right)^2 \mathrm{d}f + \int_B \left\langle \left((\mathbb{P}^{(p)}S)(f) - \hat{S}^{(p)}(f)\right)^2 \right\rangle \mathrm{d}f. \quad (9)$$

The first term in this inequality is error across B between the true PSD and its interpolating function with order p piecewise Lagrange polynomials. There are many results on the size of this error, but loosely speaking, if $P+1$ derivatives of $S(f)$ exist on all of B and B is discretized uniformly with subintervals of size ΔF, then the error between S and $\mathbb{P}^{(p)}S$ is bounded by a constant multiple of $(\Delta F)^p$.

The size of discretization and order of interpolating polynomial play a part in determining the overall MSE, along with an estimate of the second term in the inequality above. It is shown in Appendix A that the second term in (9) can be further decomposed by

$$\sum_{k=0}^{K-1}\left(\sum_{p=0}^{P}\mathcal{C}_k^{(p)}\mathbb{V}\mathrm{ar}\left(c_k^{(p)}\right) + \sum_{q\neq p}\mathcal{C}_k^{(pq)}\mathbb{C}\mathrm{ov}\left(c_k^{(p)}, c_k^{(q)}\right)\right) + \sum_{k=0}^{K-1}\sum_{p=0}^{P}\mathcal{C}_k^{(p)}\left\langle S(f_k^{(p)}) - c_k^{(p)}\right\rangle^2, \quad (10)$$

where symbols $\mathbb{V}\mathrm{ar}$ and $\mathbb{C}\mathrm{ov}$ are variance and covariance operations, respectively. The variables $\mathcal{C}_k^{(p)}$ and $\mathcal{C}_k^{(pq)}$ are defined in Appendix A. It will be clear that the first and second terms in (10) indicate the variance and the bias of the PSD estimate, respectively. Further expansion of these terms for a specific example is detailed in Section III-D.

Examining (9) and (10) reveals an interplay between the order of interploation, size of ΔF, variance and covariance estimates of the unknown coefficients $c_k^{(p)}$, and the bias of $c_k^{(p)}$ described by $\left\langle S(f_k^{(p)}) - c_k^{(p)}\right\rangle$.

In the following, we show how this mean square estimate falls within the typical multitaper PSD estimators introduced in [18]. For these estimates, we will consider only the piecewise constant interpolants typically studied in PSD estimation. This simplifies the problem because $P=0$ and the covariance term in (10) drops out.

We would like to reiterate that, to our best knowledge, the generalized MSE formulations in (9) and (10) are new results. The published literature mostly dismissed the first term of (9) and only contained discussion in respect to the second term. The formulation we discuss here provides more flexibility in the assumptions which can be imposed on the PSD estimate and includes extra error terms that comes from the added flexibility. In addition, most studies considered a flat spectrum in each B_k, equivalently a zero-order polynomial interpolation. The expression in (10) provides a more general way to account for the MSE by considering the higher-order interpolation. This is particularly important if the PSD as a function of frequency is differentiable up to an order known to us. Although we will also limit our discussion to the condition $P=0$ in the remainder of the paper, we suspect that this new tool enables studying the problem in more generality than has been previously done.

C. Quadratic Estimators

At this point, we have somewhat taken care of the first part of (9) in Section III-B, where an interpretation of the PSD is made with limited to no knowledge of the signal. Now we turn focus to the second part of (9) (or equivalently (10)), in particular the estimated PSD $\hat{S}(f)$ or $c_k^{(p)}$ once the interpolant functions $\phi_k^{(p)}$ are chosen.

We start by generalizing the spectrum analysis problem from the linear-algebra perspective. As described above, the goal is to determine coefficients $c_k^{(p)}$ for each subinterval $B_k \subset B$ up to polynomial interpolant order P. We estimate these coefficients with quadratic functions of measured data, $\mathbf{z}^\dagger(\bar{f})\mathbf{Q}\mathbf{z}(\bar{f})$, where $\mathbf{Q}$ is an $N\times N$ matrix with its rank M typically bounded in the range from 1 to N. The vector variable $\mathbf{z}$ is the modulated time-series data $z_n = x_n e^{-2\pi f n}$ at the frequency of interest f and $\cdot^\dagger$ is the matrix conjugate transpose operator. Since we will only consider piecewise constant estimators (*i.e.* $P=0$), we only need to estimate $\hat{S}(f)$ at one frequency per subinterval. In other words, $\bar{f} = f_k$. Here we also suppress the superscript in $f_k^{(0)}$ for brevity.

Quadratic estimators of this type have been studied from the perspective of maximum likelihood estimation [19], low-rank linear projections [20], and linear integral equations [18]. Quadratic estimators are dimensionally correct and generalize a number of commonly studied PSD estimation techniques, as we describe below [21].

A typical assumption about $\mathbf{Q}$ is that it is positive semidefinite, *i.e.*, that $\mathbf{z}^\dagger\mathbf{Q}\mathbf{z} \geq 0$ for all $\mathbf{z}\in\mathbb{C}^N$, where $\mathbb{C}^N$ indicates N-dimensional complex vectors. This results in PSD estimates which are non-negative, a desirable property since $S(\bar{f}) \geq 0$. Under such an assumption, $\mathbf{Q}$ can be factored as $\mathbf{Q} = \mathbf{W}\mathbf{W}^\dagger$ for a positive semidefinite matrix $\mathbf{W}\in\mathbb{C}^{N\times M}$, where $\mathbb{C}^{N\times M}$ indicates $N\times M$-dimensional complex matrices. It follows immediately that

$$\mathbf{z}^\dagger\mathbf{Q}\mathbf{z} = \|\mathbf{W}^\dagger\mathbf{z}\|^2 = \mathrm{tr}\left(\mathbf{W}^\dagger\mathbf{z}\mathbf{z}^\dagger\mathbf{W}\right), \quad (11)$$

where tr(.) denotes the trace of a matrix.

1) Single-Taper Estimate: From the quadratic estimate representation in (11), we see that the tapered periodograms are quadratic estimators with the matrix $\mathbf{Q}$ of rank 1. As such, $\mathbf{Q}$ can be factored by $\mathbf{w}\mathbf{w}^\dagger$. Alternatively, $\hat{S}_\mathrm{w}(f) = \|\mathbf{w}^\dagger\mathbf{z}(f)\|^2$. For the standard periodogram, the elements of the taper vector $\mathbf{w}$ is further reduced to $w_n = 1/\sqrt{N}$. In other words, the standard periodogram uses normalized equal weights. Examples of a taper of non-equal weights are Bartlett windows, Hann windows, Blackman windows, etc.

2) Multi-Taper Estimate: In general, the factor of the matrix $\mathbf{Q}$ is composed of M vectors $[\mathbf{w}_1\mathbf{w}_2\cdots\mathbf{w}_M]$. Preferably, orthogonality among the M taper vectors are enforced; $\mathbf{w}_m^\dagger\mathbf{w}_{m'} = 0$ for $m\neq m'$. This general case results in the multi-taper estimate, which is essentially an average of M

single-taper estimates

$$\hat{S}_{\mathrm{MT}}(f) = \frac{1}{M}\sum_{m=0}^{M-1}\hat{S}_{\mathrm{w}_m}(f) = \frac{1}{M}\sum_{m=0}^{M-1}\|\mathbf{w}_m^\dagger \mathbf{z}(f)\|^2. \quad (12)$$

Each single-taper estimate is equally weighted in (12), although non-equal weight can be implemented in some cases [18], [21].

D. Statistical Properties of Quadratic Optimization

To further motivate quadratic estimation, consider the problem of estimating $S(f_k)$ for some fixed frequency $f_k \in B = \left[-\frac{1}{2}, \frac{1}{2}\right]$. We want a matrix $\mathbf{W}$ so that $\frac{1}{M}\|\mathbf{W}^\dagger \mathbf{z}(f_k)\|^2$ as close to $S(f_k)$ as possible, in expectation. In other words, we study the behavior of

$$\left\langle \frac{1}{M}\|\mathbf{W}^\dagger \mathbf{z}(f_k)\|^2 \right\rangle = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{M}\sum_{m=0}^{M-1}|\tilde{\mathbf{w}}_m(f_k - f)|^2 S(f)\,\mathrm{d}f, \quad (13)$$

where $\tilde{\mathbf{w}}_m(f_k - f)$ is a discrete Fourier transform of $\mathbf{w}_m$ at the frequency $f_k - f$. If we could pick $\mathbf{W}$ so that $\frac{1}{M}\sum_{m=0}^{M-1}|\tilde{\mathbf{w}}_m(f_k - f)|^2 = \delta(f - f_k)$, we would have a perfect estimate of $S(f_k)$. This is unfortunately not possible with finite amounts of measured data ($N < \infty$). Quadratic estimators are instead used to estimate $S(f)$ with resolution ΔF on non-overlapping subintervals $B_k \subset B$ of length ΔF, as discussed in the previous section.

Here, it's worth noting that the term $\psi(f_k - f) = \frac{1}{M}\sum_{m=0}^{M-1}|\tilde{\mathbf{w}}_m(f_k - f)|^2$ is the frequency-domain band-pass filter response associated with the windows $\mathbf{W}$. We use quadratic estimators of these quantities to minimize the bias and variance properties described in expression of (10). To determine closed-form solutions in the variance minimization, we assume the auto-correlation of the signal can be characterized by a matrix $\mathbf{R}$. Equivalently, $\mathbf{R} \equiv \langle \mathbf{x}\mathbf{x}^\dagger \rangle$. Notably, the auto-correlation matrix relates to the signal power spectrum as follows.

$$\mathbf{R} = \int_B S(f)\mathbf{e}(f)\mathbf{e}^\top(-f)\,\mathrm{d}f, \quad (14)$$

where the modulation vector $\mathbf{e}(f)$ is $[1 \;\; \exp(2\pi i f) \;\; \cdots \;\; \exp(2(N-1)\pi i f)]^\top$. Here $\cdot^\top$ is the matrix transpose operator.

The variance and bias of estimated quadratic PSDs have been studied extensively and recent non-asymptotic bounds were derived in [20]. We build on the results from [20] to find that the estimated variance and bias can be bounded by

$$\frac{1}{M^2}\left(\sqrt{N}S_{\mathrm{rsq}}^{(k)} + \sqrt{\sum_{m=0}^{M-1}\left(1-\lambda_m(W)\right)^2 S_{\max}^{(\tilde{k})}}\right)^2, \quad (15a)$$

and

$$\underbrace{\left|\bar{S}^{(k)} - \frac{S_{\min}^{(k)}}{M}\sum_{m=0}^{M-1}\lambda_m(W)\right|^2}_{\text{Local bias}} + \underbrace{\left|S_{\max}^{(\tilde{k})}\left(1 - \frac{1}{M}\sum_{m=0}^{M-1}\lambda_m(W)\right)\right|^2}_{\text{Broadband bias}}, \quad (15b)$$

respectively. A detailed derivation can be found in Appendix B. Here the parameters associated with the spectrum are

$$S_{\max}^{(\tilde{k})} = \max_{f \in B \setminus B_k} S(f), \quad (16a)$$

$$S_{\mathrm{rsq}}^{(k)} = \sqrt{\int_{B_k} S^2(f)\,\mathrm{d}f}, \quad (16b)$$

$$\bar{S}^{(k)} = \frac{1}{\Delta F}\int_{B_k} S(f)\,\mathrm{d}f, \quad (16c)$$

$$S_{\min}^{(k)} = \min_{f \in B_k} S(f). \quad (16d)$$

Notably, $\bar{S}^{(k)}$ is the zeroth order polynomial interpolant of the true spectrum $S(f)$ in the subinterval B_k. As shown in (15b), the bias is bisected into two parts; the local bias and the broadband bias. Each of them depends on the window functions. The local bias term also relates to the choice of the basis function.

In addition, $\lambda_m(W)$ are the eigenvalues of the $M \times M$ matrix $\mathbf{W}^\dagger\mathbf{E}\mathbf{W}$ dependent on the window $\mathbf{W}$. The matrix $\mathbf{E}$, also known as the prolate matrix, is the integration of $\mathbf{e}(f)\mathbf{e}^\top(-f)$ over the interval B_k. Its element $E_{nn'}$ can be represented by

$$E_{nn'} = \begin{cases} \frac{\sin\left((n-n')\pi\Delta F\right)}{(n-n')\pi}, & \text{if } n \neq n' \\ \Delta F, & \text{if } n = n'. \end{cases} \quad (17)$$

It's also worth noting that the integration $\mathbf{e}(f)\mathbf{e}^\top(-f)$ over the interval B results in an identity matrix $\mathbf{I}$.

E. Choices of Windows

We now reach the point of choosing an optimized window. From the discussion in Section III-D, we can summarize as follows for the window optimization.

1) Select an interpolant order to appropriately encode *a priori* information about S.
2) Minimize variance in (15a) by choosing proper $\mathbf{W}$.
3) Minimize both the local bias and the broadband bias in (15b) by choosing proper $\mathbf{W}$.

As discussed above, the normalized indicator function, $\chi_{B_k}(f)$, is one of the most common choices for the basis function. The indicator function is zero outside of B_k and $1/\sqrt{\Delta F}$ on B_k. If we follow the common convention by choosing the indicator function, the optimization criteria become

1) $\arg\min_W \frac{1}{M}\sqrt{\sum_{m=0}^{M-1}\left(1-\lambda_m(W)\right)^2}$.

2) $\arg\max_{W} \frac{1}{M}\sum_{m=0}^{M-1}\lambda_m(W)$.

Coincidentally, the window composed of the Slepian sequences serve the purpose to meet these criteria [16]. The Slepian sequences correspond to the eigenvectors of the prolate matrix $\mathbf{E}$. The eigenvalues of the prolate matrix has a unique property. All the eigenvalues are bounded in the range of $(0, 1)$. If they are arranged in a descending order, the first about $N\Delta F$ eigenvalues are close to 1 and the rest of them are close to 0. Recent study shows that the transition range of $\epsilon < \lambda < 1-\epsilon$ for some small ϵ grows logarithmically with $N\Delta F$ and ϵ. The closed form of the prolate, and non-asymptotic form of its eigenvalues, significantly simplify our analysis of the bias and variance of quadratic estimators. To our knowledge, similar results do not exist for the problem with higher-order interpolants. As such, we defer their study to future work.

Conventionally, the number of the sequence M is chosen to match the number of close to 1 eigenvalues. That is $M = N\Delta F$. However, it is advantageous to only use the first $N\Delta F+1-\mathcal{O}(\log(N\Delta F/2))$ eigenvalues [20], particularly for large $N\Delta F$ or signals of a highly dynamic spectrum with large $S_{\max}^{(\tilde{k})}$. As it is indicated in (15a) and (15b), both the second term of the variance and the broadband bias are minimally influenced by the spectrum outside of B_k if λ_m all selected as close as to 1.

If we select the indicator function as the basis and use the window constructed by an optimal number of the Slepian sequences, the bounds of the standard deviation (square root of variance) and the bias can be approximated by

$$\begin{aligned}\mathrm{Std}(\hat{S}(f_k)) &\approx \frac{\sqrt{N}}{M}S_{\mathrm{rsq}}^{(k)} + \delta S_{\max}^{(\tilde{k})} \\ &\leq \frac{1}{\sqrt{M}}\hat{S}(f_k) + \delta S_{\max}^{(\tilde{k})}, \end{aligned} \tag{18a}$$

$$\text{Local bias} \leq \bar{S}^{(k)} - S_{\min}^{(k)} + \delta S_{\min}^{(k)}, \tag{18b}$$

$$\text{Broadband bias} \leq \delta S_{\max}^{(\tilde{k})}, \tag{18c}$$

where δ is the difference between the 1 and the eigenvalue λ_{M-1}, the smallest among M eigenvalues. In addition, the fact $M \sim \mathcal{O}(N\Delta F)$ is used for the upper bound in (18a). Loosely speaking, the symbol $\mathcal{O}(\cdot)$ means on-the-order-of.

Evidently, the choice of multiple tapers consisting of the first M Slepian sequences not only minimizes both the local and the broadband bias but also reduces the variance. Use of the single-taper estimate would results in the standard deviation of the PSD as large as the nominal value of the PSD, namely $\sigma_{\hat{S}(f)} \approx \hat{S}(f)$, which is consistent with conventional results. The standard deviation is scaled inversely with the squared root of the number of windows, M. However, it becomes a tradeoff between the first and the second term in (18a) once $1-\lambda_m$ is no longer negligibly small or the PSD out of the local frequency band is comparably strong.

IV. Bias Quantification for General Windows

Although the Slepian-sequence constructed window provides an estimate with the minimum MSE, there are certain cases when other window functions are favored, especially when the computational expense is a concern. Once a general window is chosen, bounds tighter than (18) can be established, particularly for the broadband bias. The idea is to replace the true PSD $S(f)$ with the estimated PSD $\hat{S}(f)$. As a result, the integration involving the frequency-dependent PSD in (27) becomes readily amenable to numerical calculation.

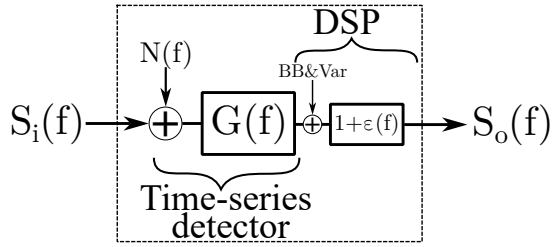


Fig. 1. A general model of the signal detection instrument followed by DSP of time-series data. The ultimate goal is to determine the input signal $S_i(f)$ from the processed signal $S_o(f)$. $G(f)$ and $N(f)$ are systematics and can be determined by system calibration with at least two known references $S_i(f)$. The local bias $(1+\epsilon)$, the broadband bias (BB) and the variance (Var) are associated with PSD estimate of a time series.

The broadband-bias integration can be approximated by

$$\begin{aligned}\frac{1}{M}\int_{B\setminus B_k} S(f')\mathrm{tr}(\mathbf{A}(f_k)\mathbf{e}(f')\mathbf{e}^\top(-f'))\,\mathrm{d}f' \\ \approx \int_{B\setminus B_k}\hat{S}(f')\psi(f'-f_k)\,\mathrm{d}f'.\end{aligned} \tag{19}$$

The convolution theorem can be applied to the integration above for implementing numerical calculation in a computer. It is straightforward to show

$$\begin{aligned}\int_{B\setminus B_k}\hat{S}(f')\psi(f'-f_k)\,\mathrm{d}f' \\ = \mathcal{F}^{-1}\left(\mathcal{F}(\hat{S})\odot\mathcal{F}(\psi)\right)_k - \hat{S}(f_k)\psi_{\max},\end{aligned} \tag{20}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the DFT and the inverse DFT operations, respectively. The symbol $\odot$ in the parenthesis indicates the point-wise product of two series. $\psi_{\max}$ is the largest filter coefficient associated with the window $\mathbf{W}$. To our best knowledge, the above equation is a new result that provides a numerical way to quantify the broadband bias from the estimated PSD.

We also want to say a few words about the local bias. Referring to (18b), the local bias can be approximated by $\epsilon_k\hat{S}(f_k)$ for signals devoid of strong variation in a subinterval B_k (*i.e.* $S_{\min}^{(k)} \approx \bar{S}^{(k)}$). Here ϵ_k is the difference between 1 and the average of M eigenvalues λ_m. ϵ_k does not have dependence on the signal PSD outside the B_k. As a result, such a bias with a small multiplicator ϵ_k can be accounted for in the instrument calibration step, as shown in Fig. 1. An instrument for signal detection can be modeled as a two port network with a transfer function $G(f)$ and the added noise $N(f)$. The DSP stage can be viewed as an additional cascaded network. Consequently, the local bias term $(1+\epsilon(f))$ can be lumped together with the system transfer function $G(f)$. The use of at least two known references at the input can resolve the overall system gain $G(f)(1+\epsilon(f))$ during the

calibration. However, the broadband bias and the variance cannot be resolved during calibration since both of them are dependent on the PSD outside of B_k. They can only be taken into account by uncertainty analysis.

Furthermore, we emphasize that the window choice discussed in previous section is for the sole purpose of minimizing the estimation error, namely the combination of the variance and the bias. In other words, this approach is applicable when the measurement accuracy is a priority. However, other practical applications may prefer measurement speed over measurement accuracy, as will be discussed in Section V. Even for applications of high-precision requirements, the instrument measurement uncertainty other than variance and bias introduced by DSP may be predominantly large. As such, the variance and bias associated with DSP become negligible in the overall uncertainty evaluation.

Nevertheless, the formulations of the variance and bias components presented in Section III-E and IV enabled us to quantify the measurement uncertainty associated with the DSP as indicated in Fig. 1. Although the origin of each component is different, both the analog and digital signal chains in a measurement system can be modeled in a similar way, namely a data-in-and-data-out two-port terminal with a transfer function. The transfer function can be determined by either instrument calibration for the hardware step or statistical analysis for the software step. The latter is what we have focused on in this paper.

V. Examples and Discussion

We apply spectrum analysis techniques and bias quantification on three of practical examples. We use both the simulated data and experimental data for PSD estimation comparison. The Welch method employing commonly known windows are often integrated in off-the-shelf signal analyzers. In the following examples, we compare the PSD estimated by the Welch method with the PSD estimated by the less conventional Slepian multi-taper method.

A. Two Continuous-Wave Tones

The first example is the time series of two continuous-wave (CW) tones generated by simulation. One tone at 50 MHz was set slightly stronger than the other tone at 125 MHz. We choose three different window types to process the data.

1) Hann window with $w_{\mathrm{H}}(k) = 0.5 + 0.5\cos\left(\frac{\pi k}{K}\right)$,
2) Blackman window with $w_{\mathrm{B}}(k) = 0.42 + 0.5\cos\left(\frac{\pi k}{K-1}\right) + 0.08\cos\left(\frac{\pi k}{K-1}\right)$,
3) Slepian multiple tapers.

The first two methods are essentially the non-overlapping Welch method with a specific window with a length of K. The time series with length of N is divided to $\lfloor N/K \rfloor$ sub-records over which all PSD estimates on the individual sub-record are averaged. The symbol $\lfloor\cdot\rfloor$ is the floor operation. This process can be equivalently viewed as a multi-taper estimate with a $N \times \lfloor N/K \rfloor$ window matrix $\mathbf{W}$. Each column vector $\mathbf{w}_m$ is composed of a shifted Hann or Blackman window and the rest entries of $\mathbf{w}_m$ are zeros. These column vectors also fulfills orthogonality requirements just like the Slepian sequences. We choose the proper number of the Slepian sequences in order to match the frequency bin size of the two non-overlap Welch methods.

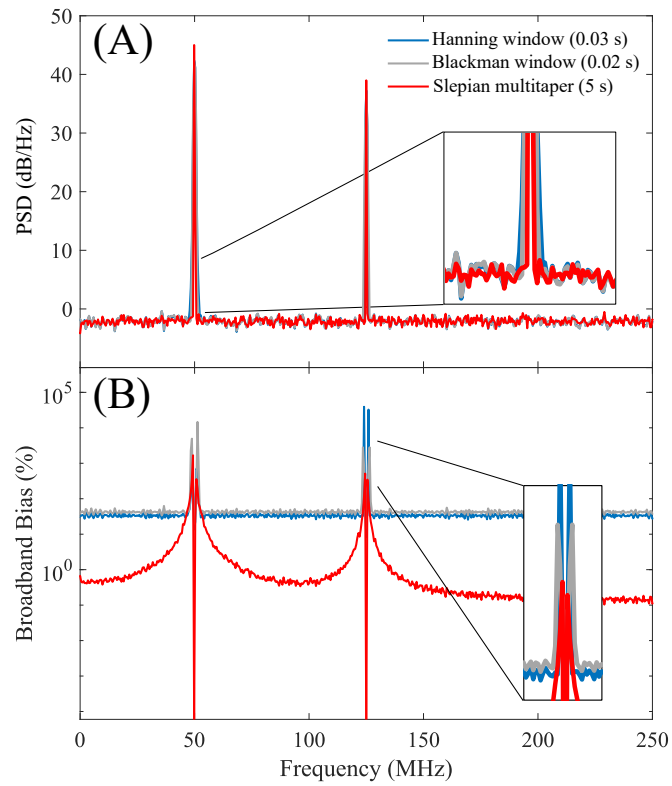


Fig. 2. (A) PSD of two CW tones at 50 MHz and 125 MHz generated by the Welch method with the Hann window and the Blackman window and the Slepian multi-taper method. The time in second in the parenthesis is the computation time for each method. (B) Comparison of estimated broadband bias of the three estimates.

The PSD estimates and the broadband bias are plotted in Fig. 2. The sample size was 2^{13} and the sampling rate was 1×10^9 Gsps. Estimates using the Hann window and the Blackman window require significantly less amount of computational time (tens of milliseconds). The multi-taper estimate with the Slepian sequences take more than two orders of magnitude of time. As shown in Fig. 2(A), two PSD peaks are identified clearly by three different methods and their relative strengths are accurately detected. The Slepian multi-taper estimate results in a sharper tone in comparison to the Hann window and Blackman window estimates.

Both the Hann window and the Blackman window estimates show a frequency-independent broadband bias at the noise floor, while the Slepian multi-taper estimate shows higher bias as the frequency gets closer to where the CW tones locate. For the noise floor far away from the tones, the broadband bias associated with Slepian estimate is about two orders of magnitude lower than that of the other two estimates. In addition, the Slepian estimate outperforms at the frequencies adjacent to the tones. Furthermore, the Slepian estimate provides significantly higher precision for the CW tone strength value as indicated by a dip of the broadband bias at the CW tone frequencies, while the two Welch estimates show about 10% broadband

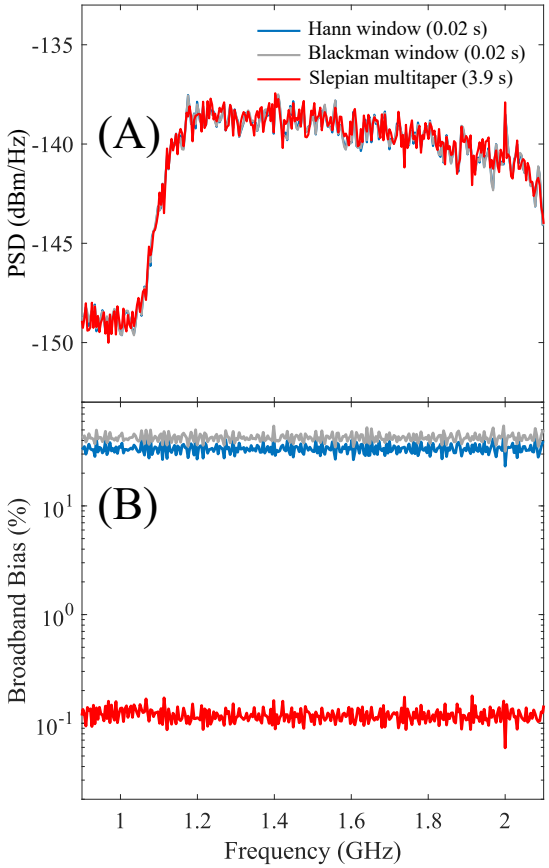


Fig. 3. (A) PSD of an amplified noise source generated by the Welch method with the Hann window and the Blackman window and the Slepian multi-taper method. The computation time in second is listed in the parenthesis for each method. (B) Comparison of broadband bias of the three estimates.

bias. The inset on the bias plot also indicates that the stronger tone has influence to the broadband bias at the frequencies in the vicinity of the weaker tone for single-taper estimates. This is especially noticeable for the Hann-windowed estimate.

B. Stationary Noise Source

The second example is a noise diode amplified by a low-noise amplifier (LNA). The sample size was 2^{13} and the sampling rate was 1×10^9 Gsps. The LNA has a lower cut-off below 1.2 GHz. An additional filtering component was also used to truncate the noise signal above 2 GHz to prevent aliasing. The resultant noise spectrum shows appreciable frequency dependence. The sampling device we used was a commercial analog-to-digital converter (ADC). The ADC has 12 bit resolution and adjustable sampling rates up to 8 Gsps. We again apply the same PSD estimate methods to the sampled noise data as those mentioned in Section V-A. Plots are shown in Fig. 3.

Once again, the Welch method by use of either the Hann window or the Blackman window is much more computationally efficient. It is more than 100 times faster than the Slepian method. The PSD estimated by three different ways show almost indiscernible difference. They all have noticeable variance at a level of more than 10%, which can be estimated by use of (18a). The broadband bias for all methods are relatively frequency independent for the entire frequency band. Not surprisingly, the Slepian multi-taper method outperforms again by a factor of more than two orders of magnitude.

C. Communication Signal Band Occupation

The third example is the spectrum occupancy measurements of communication signals. As the consumer wireless applications continue to proliferate, the communication spectrum has become more limited than ever. Opportunistic spectrum access is one of the popular ways to efficiently manage the spectral resources for communication [22][23].

In combination with white noise, three orthogonal frequency-division multiplexing (OFDM) signals were generated in adjacent frequency bands. Two of them have a 100 MHz frequency bandwidth and the other one has 50 MHz frequency bandwidth. They are offset from the carrier frequency by 100 MHz, 200 MHz and 300 MHz, respectively. The signal-to-noise-ratio (SNR) of two stronger signals are about 60 dB (labeled as OFDM1 in Fig. 4) and 30 dB (labeled as OFDM3 in Fig. 4), respectively. The other weaker signal has a SNR of about 1 dB (labeled as OFDM2 in Fig. 4). The sample size was 2^{13} and the sampling rate was 1×10^9 Gsps. We again applied three window functions to process the discretely sampled signal.

Qualitatively speaking, the two stronger OFDM signals, OFDM1 and OFDM3, were resolved equally well by three different window types. The much weaker OFDM2 signal were somewhat disguised by the emissions in adjacent frequency bands. The PSD determined by using the Slepian multi-taper method showed the band edges of OFDM2. Although the frequency band edges were not detected by using the Hann and Blackman windows, the resolved in-band PSD levels within the OFDM2 channel were appreciably higher than the noise floor at about -90 dB. As a result, the band occupancy can still pass the binary hypothesis test by setting an appropriate threshold level [23]. The bias shown in Fig. 4 became insignificant in this application. In regard to detection speed, the PSD determination by use of Hann and Blackman windows were significantly faster. Since the spectrum access is highly dynamic, the detection speed for spectral management is of high priority.

D. Discussion

The practical examples provide us some insights to the windowed PSD estimates. We summarize them below.

1) Regardless of what windowed estimations are used, the traditional non-overlap Welch method and the Slepian method can all be considered as a multi-taper estimate. The window matrix is composed of orthogonal columns.
2) The Welch method is computationally efficient. Aside from more well known windows (Barlett, Blackman, Hamming, Hann, Kaiser, etc), some high-performance window functions developed with specific characteristics are documented in [24]. These provide users more choices than the limited number of window types available in a signal analyzer.

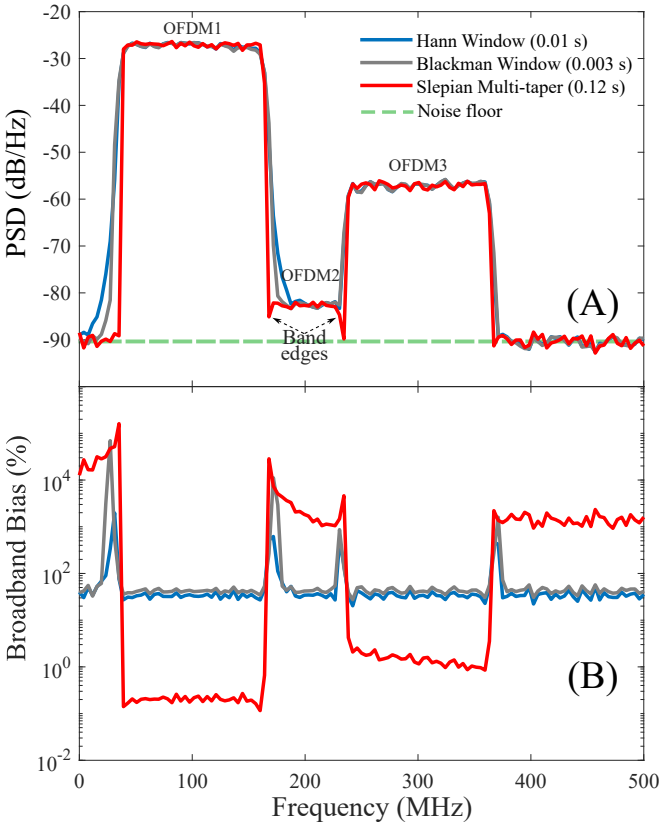


Fig. 4. (A) PSD of OFDM signals occupying three adjacent frequency bands. Processing speeds can be compared by computation times for three different methods. Although band edges of OFDM2 signal can only be resolved by Slepian multi-taper spectral estimate, OFDM2 SPDs were comparably higher than the noise floor (marked in green) by all three estimates. (B) Comparison of estimated broadband bias of the three estimates.

3) The PSD estimates provided by the Slepian sequences present challenges for implementing fast spectrum detections, although fast algorithms exist for improved efficiency especially for a large sample size [25]. The fast Slepian algorithm also handles the spectrum of a high dynamic range very well.
4) For applications of tone detections, both the Welch and the Slepian methods are sufficient. The Slepian method provides much more precise measurement of the tone strength.
5) For applications of noise measurements, the broadband bias by the Slepian method is much smaller, although all of them suffer a high variance.
6) For applications of spectrum occupancy measurements, the variance and the bias of the detected PSD are much less important. Although the Slepian method can resolve edges of occupied frequency bands even for very weak communication signals, all three methods can distinguish band occupancy by comparing in-band PSD levels with the noise floor.
7) The variance reduction becomes limited by just choosing optimal windows to minimize the broadband bias, as shown in (18a). It would require an increased number of orthogonal windows to further decrease the variance. For example, M needs to be about 10^6 for the standard deviation of estimated PSD at about 0.1%, representing ten times below the instrument uncertainty [13]. Unfortunately, this means a broadened sub-interval ΔF (equivalently a reduced frequency resolution). For absolute measurements of "pink" noise signals (frequency-dependent noise), a long time series with large N is needed for both amplitude precision and frequency resolution requirements.
8) For digital spectrum measurements, collected by sampling in the time domain, uncertainty analysis needs to include contributions from both instrument calibration and DSP. It is critical to acquire quantified broadband bias and standard deviation associated with spectrum estimation by processing the time series. These quantities can be compared to uncertainties arising from instrument calibration to assess their significance.

VI. Conclusion

This paper focuses on spectrum measurement by processing time-series data. Different from analog detectors implemented directly for measurements in the frequency domain, instruments, such as a modern signal analyzer and a digital radiometer, rely on time-domain sampling of signals within a limited frequency bandwidth and post processing the time series to generate the PSD in frequency domain. We show that the spectrum analysis problem can be generalized by quadratic estimation. The periodogram and other non-equal weight windowed estimators are all special variants of quadratic estimations.

To that end, optimization techniques for minimizing the MSE of the estimates can be used for handling the spectrum analysis. In addition, the unknown spectrum also needs to be approximately projected onto a subspace defined by a user. Such a projection may have a direct impact to the determination of the optimized windows. For a commonly used subspace spanned by the indicator functions, the Slepian-sequence constructed window can be an optimal choice to fulfill the optimization criteria. This optimal choice depends on the length of the time series N and the resolution bandwidth ΔF. Furthermore, we also present a numerical way to calculate the broadband bias (commonly known as the spectral leakage) once a window function is chosen for spectral analysis.

We show three representative examples of processing spectrum time-domain data. The window choice is ultimately determined by the application. The Welch method with commonly used windows provides efficient processing, while the Slepian method has a very low spectral leakage. The variance reduction becomes a dominant factor for noise metrology with strict requirements on the PSD amplitude precision. As a result, we would need a shorter window function for the Welch method or a larger number of sequences for the Slepian method in sacrifice of a reduction in the frequency resolution.

Appendix A
Formulations of Mean-Squared Error

We start by introducing the Lagrangely interpolated PSD into the MSE calculation between S and $\hat{S}^{(p)}$ as follows.

Triangle inequality is invoked to reach the upper bound.

$$\begin{aligned}
&\left\langle \int_B \left(S(f) - \hat{S}^{(p)}(f)\right)^2 \mathrm{d}f \right\rangle \\
&= \left\langle \int_B \Big(S(f) - (\mathbb{P}^{(p)}S)(f) \right. \\
&\qquad \left. + (\mathbb{P}^{(p)}S)(f) - \hat{S}^{(p)}(f)\Big)^2 \mathrm{d}f \right\rangle \\
&\leq \int_B \left(S(f) - (\mathbb{P}^{(p)}S)(f)\right)^2 \mathrm{d}f \\
&\qquad + \int_B \left\langle \left((\mathbb{P}^{(p)}S)(f) - \hat{S}^{(p)}(f)\right)^2 \right\rangle \mathrm{d}f. \quad (21)
\end{aligned}$$

Note that the expectation operator $\langle\cdot\rangle$ only applies to the term involving statistical variables, such as $\hat{S}^{(p)}(f)$.

The second term in (21) can be decomposed through variance properties as

$$\begin{aligned}
&\int_B \left\langle \left((\mathbb{P}^{(p)}S)(f) - \hat{S}^{(p)}(f)\right)^2 \right\rangle \mathrm{d}f \\
&\qquad = \int_B \mathbb{V}\mathrm{ar}\left(\left(\hat{S}^{(p)}(f)\right)\right) \mathrm{d}f \\
&\qquad\quad + \int_B \left\langle (\mathbb{P}^{(p)}S)(f) - \hat{S}^{(p)}(f) \right\rangle^2 \mathrm{d}f. \quad (22)
\end{aligned}$$

Expansion of $\hat{S}^{(p)}(f)$ by polynomial interpolation in (7) in combination of application of the orthgonality of $\varphi_k^{(p)}(f)$ and $\varphi_{k'}^{(q)}(f)$ for each $k \neq k'$ and $p, q = 0, \ldots, P$ allow us to derive a general expected MSE bound for the interpolated PSD estimates. In particular,

$$\begin{aligned}
&\int_B \mathbb{V}\mathrm{ar}\left(\hat{S}^{(p)}(f)\right) \mathrm{d}f \\
&\qquad = \sum_{k=0}^{K-1} \left(\sum_{p=0}^{P} \mathcal{C}_k^{(p)} \mathbb{V}\mathrm{ar}\left(c_k^{(p)}\right) \right. \\
&\qquad\qquad \left. + \sum_{q \neq p} \mathcal{C}_k^{(pq)} \mathbb{C}\mathrm{ov}\left(c_k^{(p)}, c_k^{(q)}\right) \right), \quad (23)
\end{aligned}$$

where coefficients $\mathcal{C}_k^{(p)}$ and $\mathcal{C}_k^{(pq)}$ are defined as $\mathcal{C}_k^{(p)} \equiv \int_B \left(\varphi_k^{(p)}(f)\right)^2 \mathrm{d}f$ and $\mathcal{C}_k^{(pq)} \equiv \int_B \varphi_k^{(p)}(f)\varphi_k^{(q)} \mathrm{d}f$.

In a similar manner, the second term on the right side of (22) can be represented by use of expansion of both (7) and (8).

$$\begin{aligned}
&\int_B \left\langle (\mathbb{P}^{(p)}S)(f) - \hat{S}^{(p)}(f) \right\rangle^2 \mathrm{d}f \\
&\qquad = \sum_{k=0}^{K-1} \sum_{p=0}^{P} \mathcal{C}_k^{(p)} \left\langle S(f_k^{(p)}) - c_k^{(p)} \right\rangle^2. \quad (24)
\end{aligned}$$

Appendix B
Variance and Bias of Multi-Taper Quadratic Estimation

In this appendix, we show derivation that leads to (15a) and (15b). We focus on the variance term in (23) as the covariance term no longer exists for $P = 0$. The variance term is represented by

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(c_k\right) &= \frac{1}{M^2} \mathbb{V}\mathrm{ar}\left(\|\mathbf{W}^\dagger \mathbf{z}(f_k)\|^2\right) \\
&= \frac{1}{M^2} \mathrm{tr}\left(\left(\mathbf{A}(f_k)\mathbf{R}\right)^2\right), \quad (25)
\end{aligned}$$

where the frequency-dependent matrix $\mathbf{A}(f) = \mathrm{diag}\left(\mathbf{e}(f)\right)\mathbf{W}\mathbf{W}^\dagger \mathrm{diag}\left(\mathbf{e}(-f)\right)$. Here notation $\mathrm{diag}(\cdot)$ is a diagonal matrix with its diagonal elements composed by the vector.

Referring to [20], it can be shown that the variance is bounded by the following expressions.

$$\begin{aligned}
&\frac{1}{M^2} \mathrm{tr}\left(\mathbf{A}(f_k)\mathbf{R}\mathbf{A}(f_k)\mathbf{R}\right) \\
&\leq \frac{1}{M^2} \left(\sqrt{N} S_{\mathrm{rsq}}^{(k)} + \sqrt{\mathrm{tr}\left(\left(\mathbf{I} - \mathbf{W}^\dagger \mathbf{E} \mathbf{W}\right)^2\right)} S_{\max}^{(\tilde{k})} \right)^2 \\
&= \frac{1}{M^2} \left(\sqrt{N} S_{\mathrm{rsq}}^{(k)} + \sqrt{\sum_{m=0}^{M-1} \left(1 - \lambda_m(W)\right)^2} S_{\max}^{(\tilde{k})} \right)^2. \quad (26)
\end{aligned}$$

Similarly, the bias term in (24) can be represented as follows.

$$\begin{aligned}
&\left|\bar{S}^{(k)} - c_k\right|^2 \\
&= \left|\bar{S}^{(k)} - \frac{1}{M}\left\langle \|\mathbf{W}^\dagger \mathbf{z}(f_k)\|^2 \right\rangle\right|^2 \\
&= \left|\bar{S}^{(k)} - \frac{1}{M}\mathrm{tr}(\mathbf{A}(f_k)\mathbf{R})\right|^2 \\
&= \left|\bar{S}^{(k)} - \frac{1}{M}\int_B S(f')\mathrm{tr}(\mathbf{A}(f_k)\mathbf{e}(f')\mathbf{e}^\top(-f'))\,\mathrm{d}f'\right|^2 \\
&\leq \underbrace{\left|\bar{S}^{(k)} - \frac{1}{M}\int_{B_k} S(f')\mathrm{tr}(\mathbf{A}(f_k)\mathbf{e}(f')\mathbf{e}^\top(-f'))\,\mathrm{d}f'\right|^2}_{\text{Local bias}} \\
&\qquad + \underbrace{\left|\frac{1}{M}\int_{B\setminus B_k} S(f')\mathrm{tr}(\mathbf{A}(f_k)\mathbf{e}(f')\mathbf{e}^\top(-f'))\,\mathrm{d}f'\right|^2}_{\text{Broadband bias}} \\
&\leq \left|\bar{S}^{(k)} - \frac{S_{\min}^{(k)}}{M}\mathrm{tr}\left(\mathbf{W}^\dagger\mathbf{E}\mathbf{W}\right)\right|^2 + \left|\frac{S_{\max}^{(\tilde{k})}}{M}\mathrm{tr}\left(\mathbf{I} - \mathbf{W}^\dagger\mathbf{E}\mathbf{W}\right)\right|^2 \\
&= \left|\bar{S}^{(k)} - \frac{S_{\min}^{(k)}}{M}\sum_{m=0}^{M-1}\lambda_m(W)\right|^2 \\
&\qquad + \left|S_{\max}^{(\tilde{k})}\left(1 - \frac{1}{M}\sum_{m=0}^{M-1}\lambda_m(W)\right)\right|^2. \quad (27)
\end{aligned}$$

Here the triangle inequality identity is used for the first inequality. The second inequality is achieved by taking the PSD $S(f)$ out of the integrals with the approximations of either the minimum in-band PSD $S_{\min}^{(k)}$ or the maximum out-of-band PSD $S_{\max}^{(\tilde{k})}$.

REFERENCES

[1] P. Broersen, "Automatic spectral analysis with time series models," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 2, pp. 211–216, 2002.

[2] A. Kamble, P. H. Ghare, V. Kumar, A. Kothari, and A. G. Keskar, "Spectral analysis of eeg signals for automatic imagined speech recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–9, 2023.

[3] B. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 867–880, 1997.

[4] J. Spetzler and J. Trampert, "Implementing spectral leakage corrections in global surface wave tomography," *Geophysical Journal International*, vol. 155, no. 2, pp. 532–538, 11 2003.

[5] M. A. Wieczorek and F. J. Simons, "Minimum-variance multitaper spectral estimation on the sphere," *Journal of Fourier Analysis and Applications*, vol. 13, no. 6, pp. 665–692, 2007.

[6] A.-J. van der Veen and S. J. Wijnholds, *Signal Processing Tools for Radio Astronomy*. New York, NY: Springer New York, 2013, pp. 421–463.

[7] X. Wang, Q. Meng, J. L. Han, W. Liu, and J. Zhang, "Flexible filter bank based on an improved weighted overlap-add algorithm for processing wide bandwidth radio astronomy signals," *Publications of the Astronomical Society of the Pacific*, vol. 127, no. 958, p. 1263, dec 2015.

[8] H. Lin and W.-Y. Shin, "Non-orthogonal random access in mimo cognitive radio networks: Beamforming, power allocation, and opportunistic transmission," *PLOS ONE*, vol. 12, no. 1, pp. 1–27, 01 2017.

[9] L. Xu, J. Li, and P. Stoica, "Target detection and parameter estimation for mimo radar systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 927–939, 2008.

[10] S. Dello Iacono, G. Di Leo, C. Liguori, and V. Paciello, "The obtainable uncertainty for the frequency evaluation of tones with different spectral analysis techniques," *Metrology*, vol. 2, no. 2, pp. 216–229, 2022. [Online]. Available: https://www.mdpi.com/2673-8244/2/2/13

[11] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Prentice Hall, 1993.

[12] X. Lu, D. Kuester, and D. Gu, "Digital radiometer for traceable spectrum sensing," in *2020 Conference on Precision Electromagnetic Measurements (CPEM)*, 2020, pp. 1–2.

[13] ——, "A prototype digital radiometer for noise metrology applications: A metrology-grade digital radiometer," *IEEE Microwave Magazine*, vol. 23, no. 5, pp. 57–66, 2022.

[14] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, and OIML, "Guide to the expression of uncertainty in measurement — part 6: Developing and using measurement models," *Joint Committee for Guides in Metrology*, 2020.

[15] Keysight. (2022) Signal analysis measurement fundamentals. [Online]. Available: https://www.keysight.com/us/en/assets/7018-06802/application-notes/5965-7009.pdf

[16] D. B. Percival and A. T. Walden, *Spectral Analysis for Univariate Time Series*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2020.

[17] S. C. Brenner and L. Scott, *The Mathematical Theory of Finite Element Methods*, ser. Texts in Applied Mathematics. Springer, 2008.

[18] D. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[19] P. Stoica, R. Moses, B. Friedlander, and T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 378–392, 1989.

[20] S. Karnik, J. Romberg, and M. A. Davenport, "Thomson's multitaper method revisited," *IEEE Transactions on Information Theory*, vol. 68, no. 7, pp. 4864–4891, 2022.

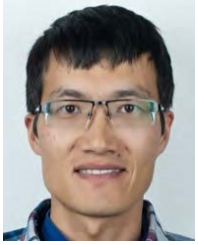
[21] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*. Cambridge University Press, 1993.

[22] M. Höyhtyä, A. Mämmelä, M. Eskola, M. Matinmikko, J. Kalliovaara, J. Ojaniemi, J. Suutala, R. Ekman, R. Bacchus, and D. Roberson, "Spectrum occupancy measurements: A survey and use of interference maps," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2386–2414, 2016.

[23] Y. Chen and H.-S. Oh, "A survey of measurement-based spectrum occupancy modeling for cognitive radios," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 848–859, 2016.

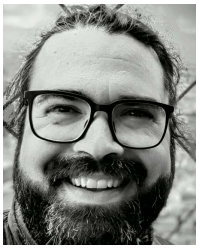
[24] G. Heinzel, A. Rudiger, and R. Schilling. (2002) Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new at-top windows. [Online]. Available: https://hdl.handle.net/11858/00-001M-0000-0013-557A-5

[25] S. Karnik, Z. Zhu, M. B. Wakin, J. Romberg, and M. A. Davenport, "The fast slepian transform," *Applied and Computational Harmonic Analysis*, vol. 46, no. 3, pp. 624–652, 2019.



Dazhen Gu received the Ph.D. degree in electrical engineering from University of Massachusetts, Amherst, in 2007.

He has been with the RF Technology Division, National Institute of Standards and Technology, Boulder, CO, since November 2003. During the first three and a half years, he did his doctoral research in development of terahertz imaging components and systems. From 2007 to 2009, he was with the Microwave Measurement Services project, where he was involved in microwave metrology, in particular thermal noise measurements and instrumentation. From 2009 to 2015, he took a position in the microwave remote-sensing project, in which a microwave brightness-temperature standard was successfully demonstrated. From 2015 to 2018, he was in charge of the microwave power project and developed the NIST power traceability with correlated uncertainties for 5G communication researches. Since March 2018, he has been with the Shared-Spectrum Metrology Group, where he is involved in the Spectrum Sensing and Noise Project. He received NIST Allen V. Astin award in 2022.



Jacob Rezac is a Mathematical Statistician with the Communications Technology Laboratory at the National Institute of Standards and Technology (NIST). He received the B.S. and M.S. degrees in applied mathematics from the Colorado School of Mines, Golden, CO, USA, in 2011 and 2012, respectively, and the Ph.D. degree in applied mathematics from the University of Delaware, Newark, DE, USA, in 2017. From 2017 to 2019, he was a Post-Doctoral Researcher with NIST and the University of Colorado at Boulder, Boulder, CO, USA. He has been with NIST since 2019. His research interests include theoretical and computational aspects of inverse problems as applied to problems in communications and the physical sciences.

Xifeng Lu is a Senior Research Fellow at University of Colorado and Research Associate at National Institute of Standards and Technology (NIST), Boulder. He received his Ph.D. in Institute of Physics, Chinese Academy of Sciences in 2005. Before moving to Colorado in 2009, he had worked at Durham University in the UK as a Postdoctoral Research Associate from 2006. He became a Senior Member of IEEE in 2011. His recent research interests include RF noise and spectrum sensing metrology, precision RF noise parameter characterization, digital radiometry, EVM estimation of millimeter-wave modulated signals, and implementing direct RF sampling in sub-Nyquist zones for the new generation of communication.

Daniel Kuester serves as project leader of the Spectrum Sensing and Noise project at NIST. His work there has focused on robust techniques for the measurement and analysis of wireless spectrum, and on metrology to connect physical waveform conditions with user-visible receiver response. In the past, he worked in industry, where he implemented design, integration, and testing of phased antenna arrays and wireless power transfer for sensors. He studied at the University of Colorado in Boulder, which awarded him undergraduate degrees in electrical engineering and music performance (2007), and the Ph.D. in electrical engineering (2012, advised by Z. Popović). His work has been recognized by the U.S. Department of Commerce with Gold and Bronze Medals and the NIST Allen V. Astin award, as well as with the "Most Innovative Use of RFID" award from *RFID Journal* and the Best Paper at the *IEEE Conference on Wireless Power Transfer*.