Dynamic Bayesian Multitaper Spectral Analysis

Proloy Das and Behtash Babadi

Abstract—Nonparametric spectral analysis using overlapping sliding windows is among the most widely used techniques in analyzing nonstationary time series. Although sliding window analysis is convenient to implement, the resulting estimates are sensitive to the window length and overlap size. In addition, it undermines the dynamics of the time series as the estimate at each window uses only the data within. Finally, estimates from overlapping windows hinder a precise statistical assessment of the estimates. In this paper, we address these shortcomings by explicitly modeling and estimating the spectral dynamics through integrating the multitaper method with state-space models in a Bayesian estimation framework. We propose two spectral estimators that are able to capture spectral dynamics at high spectrotemporal resolution. We provide theoretical analysis of the bias-variance trade-offs, which establish performance gains over the multitaper method. We apply our algorithms to synthetic data as well as real data from human EEG and electric network frequency recordings, which corroborate our theoretical analysis.

Index Terms—Nonparametric spectral analysis, state-space models, Bayesian filtering, Expectation-Maximization

I. INTRODUCTION

Spectral analysis techniques are among the most important tools for extracting information form time series data recorded form naturally occurring processes. Examples include speech [1], images [2], electroencephalography (EEG) [3], oceanography [4], climatic time series [5], and seismic data [6]. Due to the exploratory nature of most of these applications, nonparametric techniques based on Fourier methods and Wavelets are among the most widely used. In particular, the multitaper method (MTM) excels among the available nonparametric techniques due to both its simplicity and its optimality in the bias-variance trade-off [7], [8].

Most existing nonparametric spectral analysis techniques assume that the time series are stationary. In many applications of interest, however, the energy of the various oscillatory components in the data exhibit dynamic behavior due to changes in the state of the underlying system. To get a meaningful spectrotemporal description of such time series, it is commonly assumed that the underlying process is pseudo-stationary, i.e., the spectrum changes slowly with time. Thereby, the so-called spectrogram analysis is obtained by using sliding windows with overlap in order to capture nonstationarity.

Although sliding window processing is widely used, it has three major drawbacks. First, given that the estimate at each window is obtained by the data within, it ignores the common dynamic trends shared across multiple windows, and thereby fails to fully capture the degree of smoothness inherent in the signal. Instead, the smoothness of the estimates is enforced and determined by the amount of overlap between adjacent windows. Second, although techniques such as the MTM are able to reject the so-called sampling noise by averaging over multiple tapers, their spectral resolution degrades when applied to data within small windows. In addition, despite mitigating sampling noise, they are not amenable to additive measurement noise. Third, the overlap between adjacent windows hinders a precise statistical assessment of the estimates, such as constructing confidence intervals due to the high dependence of estimates across windows. To address this issue, statistical corrections for multiple comparisons need to be employed [9], which in turn limit the resulting test powers when multiple windows are involved.

In recent years, several alternative approaches to nonstationary spectral analysis have been proposed, such as the Empirical Mode Decomposition (EMD) [10], synchrosqueezed wavelet transform [11], [12], time-frequency reassignment [13], and spectrottemporal pursuit [14]. These techniques aim at decomposing the data into a small number of smooth oscillatory components in order to produce spectral representations that are smooth in time but sparse in frequency. Although they produce spectral estimates that are highly localized in the time-frequency plane, they require certain assumptions on the data to hold. For example, estimates from EMD do not take into account the random nature of the underlying signal as well as observation noise. In addition, they lack a statistical characterization of the estimates. The other methods assume that the underlying spectrottemporal components pertain to certain structures such as amplitude-modulated narrowband mixtures [11], [12], sparsity [14] or chirp-like dynamics [13]. Finally, although these sophisticated methods provide spectrotemporal resolution improvements, they do not yield implementations as simple as those of the sliding window-based spectral estimators.

In this paper, we address the above-mentioned shortcomings by integrating techniques from multitaper analysis and state-space modeling in a Bayesian estimation framework. In particular, we introduce two nonparametric estimators designed to respectively mitigate the measurement and sampling noise, while capturing the spectrottemporal dynamics of the signal and admitting fast implementations. To this end, we construct state-space models in which the underlying states pertain to certain spectral features of the data, such as the spectral eigen-coefficients or eigen-spectra in a multitaper setting. We employ state dynamics that capture the spectrottemporal evolution of the signal, coupled with observation models that capture the effect of measurement and sampling noise. We then employ Expectation-Maximization to find the maximum a posteriori (MAP) estimate of the states given the observed data to construct our spectral estimators as well as statistical confidence intervals.

The authors are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742.

This material is based upon work supported by the National Science Foundation under Grant No. 1552946.
We provide theoretical analysis of the bias-variance tradeoff, which reveals that our proposed framework inherits the optimal properties of the multitaper framework, as well as the optimal data combining and denoising features of Bayesian filtering and smoothing algorithms. In addition, due to the simplicity and wide usage of Bayesian filtering and smoothing algorithms, our algorithms are nearly as simple to implement as the sliding window-based spectrograms. To further demonstrate the performance of our algorithms, we apply them on synthetic as well as real data including human EEG during sleep and electric network frequency data from audio recordings. Application of our proposed estimators to these data provides specroteamal features that are significantly denoised, are smooth in time, and enjoy high spectral resolution, thereby corroborating our theoretical results.

The rest of the paper is organized as follows: In Section II we present the preliminaries and problem formulation. In Section III we develop our proposed estimators. Application of our estimators to synthetic and real data are given in Section IV, followed by our theoretical analysis in Section V. Finally, our concluding remarks are given in Section VI. Finally, our concluding remarks are given in Section VI.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. Nonstationary Processes and Dynamic Spectral Density

Consider a finite realization of \( T \) samples from a discrete-time nonstationary process \( y_t, t = 1, 2, \cdots T \), obtained via sampling a noise-corrupted continuous-time signal above Nyquist rate. We assume that the nonstationary process \( y_t \) is harmonizable so that it admits a Cramér representation \([15]\) of the form:

\[
y_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi f t} dz(f),
\]

where \( dz(f) \) is the generalized Fourier transform of the process. This process has a covariance function of the form:

\[
\Gamma(\tau, t) := \mathbb{E}[y_{t+\frac{\tau}{2}} y_{t+\frac{1}{2}}] = \frac{1}{2} \int_{-1/2}^{1/2} e^{i2\pi f (t+\frac{\tau}{2})} \gamma(f, g) df dg.
\]

where \( f \) and \( g \) are the stationary and nonstationary frequencies, respectively, and \( \gamma(f, g) df dg := \mathbb{E}[dz(f + \frac{\tau}{2}) dz^*(f - \frac{\tau}{2})] \) is the generalized spectral density. To avoid possible \( \delta \)-function singularity at \( g = 0 \), one can instead work with the expected value of Wigner distribution \([10]\) defined as:

\[
D(t, f) := \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi f \tau} \gamma(g, f) df dg = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi f \tau} \mathbb{E}[y_{t+\frac{\tau}{2}} y_{t+\frac{1}{2}}] d\tau.
\]

Note that \( D(t, f) \) captures the spectral representation of the data as a function of time, and provides a useful device for analyzing nonstationary time series. However, estimating \( D(t, f) \) from finite samples of the process is not straightforward. A naïve way of estimating this dynamic spectrum is to subdivide the data into windows or segments of length \( W \) samples and find the spectral representation for each window independently, under the assumption of local second-order stationarity \([17]\). This representation is popularly known as ‘Spectrogram’. The spectral representation of each data segment can be computed using various Fourier or wavelet-based methods, depending on the problem requirements. Among these methods, the Multi taper Method (MTM) is well-practiced and moreover well-studied from a statistical viewpoint.

B. The MTM Spectral Analysis

The MTM is an extension of single-taper spectral analysis, where the data is element-wise multiplied by a taper prior to forming the spectral representation to mitigate spectral leakage \([7], [18]\). In the MTM, spectral representation is computed as the average of several such single-taper PSDs, where the tapers are orthogonal to each other. This can be achieved by using the Discrete Prolate Spheroidal sequences (dpss) or Sepliam sequences \([19]\), due to their optimal leakage properties.

Another viewpoint of the MTM with this particular choice of data tapers is the decomposition of the spectral representation of the process over a set of orthogonal basis functions. Indeed, these basis functions originate from an approximate solution to the integral equation connecting the Fourier transform of the data to the spectrum:

\[
y(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sin \left( \frac{N}{2} \pi f \right) e^{-i2\pi (f - \zeta)} \frac{\delta(f - \zeta)}{\sin \pi (f - \zeta)} \, dz(\zeta),
\]

where \( W \) is window length, i.e., number of samples, and \( dz(\zeta) \) is an orthogonal increment process. This integral equation can be approximated using a local expansion of the increment process over an interval \([B, B] \), for some small design bandwidth \( B \), in the space spanned by the eigenfunctions of the Dirichlet kernel \( \frac{\sin W f}{\sin \pi f} \). \([7], [18]\).

These eigenfunctions are known as the Prolate Spheroidal Wave Functions (PSWFs), which are a set of doubly-orthogonal functions over \([B, B] \) and \([1/2, 1/2] \), with time-domain representations given by the dpss sequences. Let \( u_n^{(k)} \) be the \( n \)th sample of the \( k \)th dpss sequence, for a given bandwidth \( B \) and window length \( W \). The \( k \)th PSWF is then defined as:

\[
U^{(k)}(f) := \sum_{n=0}^{W-1} u_n^{(k)} e^{-i2\pi f n}.
\]

Using these set of data tapers, the multitaper spectral representation can be calculated as follows:

\[
\hat{S}^{(mtm)}(f) := \frac{1}{K} \sum_{k=1}^{K} |x^{(k)}(f)|^2,
\]

where \( x^{(k)}(f) := \sum_{n=0}^{W-1} e^{-i2\pi f n} u_n^{(k)} y_n \) for \( k = 1, 2, \cdots , K \) are called the ‘eigen-coefficients’. The ‘eigen-spectra’ defined as \( \hat{S}^{(k)}(f) := |x^{(k)}(f)|^2 \) can be viewed as the expansion coefficients of the decomposition.

To handle dynamic spectra under the multitaper framework, sliding windows with overlap are often used to enforce temporal smoothness and increase robustness of the estimates \([8], [17]\). Although this ‘overlapped’ MTM procedure overcomes frequency leakage issues and produces consistent estimates, this method’s heavy reliance on the degree of overlap for temporal smoothness, as opposed to capturing the evolution of the nonstationary process, makes it an ad hoc procedure. In addition, these estimates hinder precise statistical inference procedures, such as hypothesis testing, due to the statistical dependence induced by the overlaps. The objective of this
work is to overcome these limitations by directly modeling and estimating the evolution of the process without resorting to overlapped sliding windows, while achieving fast and efficient implementations.

C. Problem Formulation

In order to address the shortcomings of the overlapped sliding window analysis and to capture the dynamics of the spectrum, we employ state-space models. To this end, we consider the unobserved spectral quantities, such as the eigen-coefficients or the true eigen-spectra of the process, as the underlying states which evolve in time and generate the data. Then, Bayesian estimation techniques can be used to estimate the unobserved states and hence form dynamic PSD estimates.

Assume, without loss of generality, that an arbitrary window of length $W$ is chosen so that for some integer $N$, $N W = T$ and let $y_n = [y_{(n-1)W+1}, y_{(n-1)W+2}, \ldots, y_{nW}]^\top$ for $n = 1, 2, \ldots, N$, denote the data in the $n$th window. This way, the entire data is divided into $N$ non-overlapping segments of length $W$ each. Motivated by the major sources of uncertainty in spectral estimation, i.e. sampling noise and measurement noise, we formulate two state-space frameworks in the following subsections.

1) Mitigating the Measurement Noise: Let $u_{(k)}$ denote the $k$th dpss taper and $v_{(k)} := u_{(k)} \odot y_n$, where $\odot$ denotes element-wise multiplication. For a frequency spacing of $2\pi/J$ with $J$ an integer, let $F_n$ be a matrix with elements $(F_n)_{l,j} := \exp(2\pi i ((n-1)W + l/2))$ for $l = 1, 2, \ldots, W$ and $j = 1, 2, \ldots, J$. Let $x^{(k)}(f)$ and $y^{(k)}(f) := [x^{(k)}(0), x^{(k)}(2\pi 1/J), \ldots, x^{(k)}(2\pi J-1)]^\top$ denote the $k$th spectral eigen-coefficient and its discretized version, respectively, for $k = 1, 2, \ldots, K$. Then, we consider the following spectrotemporal representation of the tapered data segments:

$$y_n = F_n x_n + v_n,$$

where $v_n$ is independent of $x_{n-1}$, and assumed to be identically distributed according to a zero-mean Gaussian distribution with covariance $\text{Cov}\{v_{(k)}, v_{(k)}\} = \sigma^2 I_{\delta_{ij}}$. We view $y^{(k)}(f)$ as a noisy observation corresponding to the true eigen-coefficient $x^{(k)}(f)$, which provides a linear Gaussian forward model for the observation process.

In order to capture the evolution of the spectrum and hence systematically enforce temporal smoothness, we impose a stochastic continuity constraint on the eigen-coefficients $\{x_{(k)}n\}_{j=1}^{N}$ for $n = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, K$, using a first-order difference equation:

$$x_{(k)n} = \alpha_{(k)} x_{(k)n-1} + w_{(k)n},$$

where $0 \leq \alpha_{(k)} < 1$, and $w_{(k)n}$ is independent of $x_{(k)n-1}$ and assumed to be independently distributed according to a zero-mean Gaussian distribution with covariance $\text{Cov}\{w_{(k)}, w_{(k)}\} = Q_{(k)} \delta_{ij}$. Under this assumption, the discrete-time process, $(x_{(k)n})_{n=1}^{N}$ forms a jointly Gaussian random process with independent increments, while the process itself is statistically dependent. An estimate of the unobserved states (true eigen-coefficients) from the observations (tapered data) under this model suppresses the measurement noise and captures the state dynamics.

2) Mitigating the Sampling Noise: The empirical eigen-spectra of a process can be thought of the true eigen-spectra corrupted by sampling noise in a multiplicative fashion, due to having access to only a single realization of the process. For now, consider only a single window of length $W$. It is known that when the spectrum does not rapidly vary over the chosen design bandwidth $B$, the eigen-spectra are approximately uncorrelated and the following approximation holds for $k = 1, 2, \ldots, K$ [13]:

$$\frac{\hat{S}(f)}{S(f)} \sim \frac{\chi_2^2}{2}, \quad 0 < f < 1/2,$$

where $\hat{S}(f)$ and $S(f)$ are the tapered estimate and true PSD, respectively. By defining $\psi_{(k)}(f) := \log \hat{S}(k)_{(f)} + \log 2$ and $s_{(k)}(f) := \log S_{(k)}(f)$, we can transform the multiplicative noise effect of Eq. (6) to the following additive forward model [20]:

$$\psi_{(k)}(f) = s_{(k)}(f) + \phi_{(k)}(f),$$

where $\phi_{(k)}(f)$ is a log-chi-square distributed random variable, capturing the uncertainty due to sampling noise. It can be shown that $\phi_{(k)}(f)$ has a density given by:

$$p(\phi) = \frac{1}{2} \exp \left( \phi - \frac{1}{2} \exp(\phi) \right),$$

which belongs to the family of log-Gamma distributions, including the Gumbel and Bramwell-Holdsworth-Pinton distributions common in extreme value statistics [21].

In order to incorporate this observation model in our dynamic framework, we define the state vector, the observation vector and the observation noise vector as follows:

$$s_{(k)}(f) = [s_{(k)}(0), s_{(k)}(2\pi 1/J), \ldots, s_{(k)}(2\pi J-1)]^\top,$$

$$\psi_{(k)}(f) = [\psi_{(k)}(0), \psi_{(k)}(2\pi 1/J), \ldots, \psi_{(k)}(2\pi J-1)]^\top,$$

$$\phi_{(k)}(f) = [\phi_{(k)}(0), \phi_{(k)}(2\pi 1/J), \ldots, \phi_{(k)}(2\pi J-1)]^\top.$$
States Data dpss tapers another MAP-estimation problem:
framework(9), the eigen-spectra can be obtained by solving
\( k \) for \( DBMT \) and (12) the Dynamic Bayesian Multitaper (DBMT)
Bayesian estimation, in which the Bayesian risk/loss function,
D. The Inverse Problem
a visual illustration of the proposed modeling paradigm.

D. The Inverse Problem
We formulate the spectral estimation problem as one of
Bayesian estimation, in which the Bayesian risk/loss function,
fully determined by the posterior density of \( (x_n^{(k)})_{n=1,k=1}^{N,K} \)
(resp. \( (s_n^{(k)})_{n=1,k=1}^{N,K} \)) given the observations \( (y_n^{(k)})_{n=1,k=1}^{N,K} \)
(resp. \( (\psi_n^{(k)})_{n=1,k=1}^{N,K} \)) is minimized. We first consider
the forward model of [4], which provides the observed data
likelihood given the states. Under the state-space model of
Eq. [5], the \( k \)th eigen-coefficient can be estimated by solving
the following maximum a posteriori (MAP) problem:
\[
\min_{x_n^{(k)}, \ldots, x_N^{(k)}} \sum_{n=1}^{N} \left[ \frac{1}{\sigma^2} \left\| y_n^{(k)} - F_n x_n^{(k)} \right\|^2_2 + \right.
\left( x_n^{(k)} - \alpha x_{n-1}^{(k)} \right)^H Q_n^{(k)} \left( x_n^{(k)} - \alpha x_{n-1}^{(k)} \right), \quad (11)
\]
for \( k = 1, 2, \ldots, K \). Similar as this, in the second state space
framework[5], the eigen-spectra can be obtained by solving
another MAP-estimation problem:
\[
\min_{s_n^{(k)}, \ldots, s_N^{(k)}} \sum_{n=1}^{N} \left[ \frac{1}{\sigma^2} \left| \psi_n^{(k)} + \frac{1}{2} \exp(\psi_n^{(k)} - s_n^{(k)}) \right|^2_2 \right.
\left. + (s_n^{(k)} - \beta s_{n-1}^{(k)})^H R_n^{(k)} (s_n^{(k)} - \beta s_{n-1}^{(k)}) \right], \quad (12)
\]
for \( k = 1, 2, \ldots, K \), where \( 1_f \) is the vector of all ones
of length \( J \). We call the MAP estimation problems in (11)
and (12) the Dynamic Bayesian Multitaper (DBMT) and the
log-DBMT estimation problems, respectively. Similarly, the
respective PSD estimates will be denoted by the DBMT and
log-DBMT PSD estimates.

Equation (11) is a strictly convex function of \( x_n^{(k)} \in \mathbb{C}^W \)
and \( Q_n^{(k)} \in \mathbb{S}^{W+}_{++} \) for \( n = 1, 2, \ldots, N \), which can be solved
using standard optimization techniques. However, these tech-
niques do not scale well with the data length \( N \). A careful
examination of the log-posterior reveals a block tri-diagonal
structure of the Hessian, which can be used to develop
efficient recursive solutions that exploit the temporal structure
of the problem. A similar argument holds for the optimization
problem in (12). However, the parameters of these state-
space models need to be estimated from the data. In the next
section, we show how the Expectation-Maximization (EM)
algorithm can be used to both estimate the parameters and
states efficiently from the optimization problems (11) and (12).

III. Fast Recursive Solutions via the EM Algorithm
In order to solve the MAP problem in (11), we need to
find the parameters \( Q_n^{(k)} \in \mathbb{S}^{W+}_{++} \) and \( \alpha^{(k)} \in (0,1) \) for
\( n = 1, 2, \ldots, N \) and \( k = 1, 2, \ldots, K \). Similarily \( R_n^{(k)} \in \mathbb{S}^{W+}_{++} \)
and \( \theta^{(k)} \in (0,1) \) need to be estimated for the problem in
(12). If the underlying states were known, one could further
maximize the log-posterior with respect to the parameters. This
observation can be formalized in the EM framework [22–24].
To avoid notational complexity, we drop the dependence of
the various variables on the taper index \( k \) in the rest of
this subsection.

A. The DBMT Spectrum Estimation Algorithm
By treating \( x_n, n = 1, 2, \ldots, N \) as the hidden variables
and \( \alpha, Q_n, n = 1, 2, \ldots, N \) as the unknown parameters to be
estimated, we can write the complete log-likelihood as:
\[
\log L(\alpha, Q_{1:N}) := -\sum_{n=1}^{N} \left[ \frac{1}{\sigma^2} \left\| y_n - F_n x_n \right\|^2_2 + \log |\det Q_n| \right.
\left. + (x_n - \alpha x_{n-1})^H Q_n^{-1}(x_n - \alpha x_{n-1}) \right] + \text{cnst.} \quad (13)
\]
For simplicity of exposition, we assume that \( Q_n = Q \) for
\( n = 1, 2, \ldots, N \). The forthcoming treatment can be extended
to the general case with little modification. Also, note that \( \sigma^2 \)
can be absorbed in \( Q \), and thus is assumed to be known. At
the 0th iteration, we have:

1) E-Step: Given \( \alpha^{[l]}, Q^{[l]} \), for \( n = 1, 2, \ldots, N \), the expectations
\( x_{n|N} := \mathbb{E}[x_n|y_{1:N}, \alpha^{[l]}, Q^{[l]}] \),
\( \Sigma_{n|N} := \mathbb{E}[(x_n - x_{n|N})(x_n - x_{n|N})^H|y_{1:N}, \alpha^{[l]}, Q^{[l]}] \),
\( \Sigma_{n,n-1|N} := \mathbb{E}[(x_n - x_{n|N})(x_{n-1} - x_{n-1|N})^H|y_{1:N}, \alpha^{[l]}, Q^{[l]}] \),
can be calculated using the Fixed Interval Smoother (FIS) [25]
and the state-space covariance smoothing algorithm [26].
These expectations can be used to compute the expectation of
the complete data log-likelihood \( \mathbb{E}[\log L(\alpha, Q)|y_{1:N}, \alpha^{[l]}, Q^{[l]}] \).

2) M-Step: The parameters \( \alpha^{[l+1]} \) and \( Q^{[l+1]} \) can be obtained
by maximizing the expectation of (13). Although this
expectation is convex in \( \alpha \) and \( Q \) individually, it is not a
convex function of both. Hence, we perform cyclic iterative
updates for \( \alpha^{[l+1]} \) and \( Q^{[l+1]} \) given by:
\[
\alpha^{[t+1]} = \frac{\sum_{n=1}^{N} \text{Tr}(\Sigma_{n-1:n}|x_{0:N}) + x_{n}^H \Sigma_{n-1:n}^{-1} x_{n|N} \text{I}}{\sum_{n=1}^{N} \text{Tr}(\Sigma_{n-1:n}|x_{0:N}) + x_{n}^H \Sigma_{n-1:n}^{-1} x_{n|N}}
\]
(14)
and
\[
Q_{k+1} = \frac{1}{N} \sum_{n=1}^{N} \left[ x_{n|N} x_{n}^H + \Sigma_{n|N} + \alpha^{[t+1]} x_{n|N} + \Sigma_{n-1|N} \right] - \alpha^{[t+1]} \left( x_{n|N} x_{n}^H + x_{n|N} x_{n}^H + 2 \Sigma_{n|N} \right).
\]
(15)

These iterations can be performed until convergence to a possibly local maximum. However, with even one such update, the overall algorithm forms a majorization-minimization (MM) procedure, generalizing the EM procedure and enjoying from similar convergence properties \cite{27}. One possible implementation of this iterative procedure is described in Algorithm 1. Once the DBMT estimates of all the \(K\) eigen-coefficients \(x_{n}^{(k)}\) are obtained, for \(n = 1, 2, \cdots, N\) and \(k = 1, 2, \cdots, K\), the DBMT spectrum estimate is constructed similar to (3):
\[
\hat{D}_{n}(f_j) = \frac{1}{K} \sum_{k=1}^{K} \left| \hat{S}_{n,j}^{(k)} \right|^2,
\]
(16)
where \(f_j := \frac{2\pi(j-1)}{N} \) for \(j = 1, 2, \cdots, J\) and \(n = 1, 2, \cdots, N\). Confidence intervals can be computed by mapping the Gaussian confidence intervals for \(x_{n}^{(k)}\) to the final DBMT estimate.

B. The log-DBMT Spectrum Estimation Algorithm

We utilize a similar iterative procedure based on the EM algorithm to find the log-DBMT spectrum estimate. As before, we treat \(s_{n}, n = 1, 2, \cdots, N\) as hidden variables and \(\theta, R_{n}, n = 1, 2, \cdots, N\) as the unknown parameters to be estimated. In order to give more flexibility to the observation model, we consider the observation noise to be distributed as log-chi-square with degrees of freedom \(2\nu\), for some positive integer \(\nu\) to be estimated. The density of each element of \(\phi_{n}^{(k)}\) is then given by:
\[
p(\phi) = \frac{1}{2\Gamma(\nu)} \exp \left( \frac{\nu\phi - 1}{2} \exp(\phi) \right).
\]
(17)

We can express the complete data log-likelihood as:
\[
\log L(\nu, \theta, R_{1:n}) := -\sum_{n=1}^{N} \left[ \frac{1}{2} \left( \nu s_{n} - \psi_{n} \right) + \frac{1}{2} \exp(\psi_{n} - s_{n}) \right] + J \left( \nu \log 2 + \log \Gamma(\nu) \right) + (s_{n} - \theta s_{n-1})^H R_{n}^{-1} (s_{n} - \theta s_{n-1}) + \log(\det R_{n}) + \text{const.}
\]
(18)
Again assuming \(R_{n} = R\) for all \(n = 1, 2, \cdots, N\) for simplicity, the following EM algorithm can be constructed:

1) E-step: Computation of the conditional expectation of the log-likelihood in (18) requires evaluating \(E[s_{n|N}, R_0, \theta|0, \nu|0] \) and \(E[\exp(-s_{n|N})|\psi_{1:n}, R_0, \theta|0, \nu|0] \) for \(n = 1, 2, \cdots, N\). Unlike the DBMT estimation problem, the forward model in this case is non-Gaussian, and hence we cannot apply the Kalman filter and FIS to find the state expectations. To compute the conditional expectation, the distribution of \(s_{n|N}, R_0, \theta|0, \nu|0\) or its samples are required \cite{28}. Computation of the distribution \(s_{n|N}, R_0, \theta|0, \nu|0\) involves intractable integrals and sampling from the distribution using numerical methods such as Metropolis-Hastings is not computationally efficient, especially for long data, given that it has to be carried out at every iteration.

Since the posterior distribution is unimodal and a deviation from the Gaussian posterior, we approximate the distribution of \(s_{n|N}, R_0, \theta|0, \nu|0\) as a Gaussian distribution by matching its mean and covariance matrix to the log-posterior in Eq. (18). To this end, the mean is approximated by the mode of \(f_{s_{n|N}|\psi_{1:n}, R_0, \theta|0, \nu|0} \) and the covariance is set to the inverse of the negative Hessian of the log-likelihood in (18) \cite{29}. Under this approximation, computing \(E[\exp(-s_{n|N})|\psi_{1:n}, R_0, \theta|0, \nu|0] \) is also facilitated thanks to the closed-form moment generating function of \(z \sim N(\mu, \Sigma)\):
\[
E[\exp(a^T z)] = \exp \left( a^T \mu + \frac{1}{2} a^T \Sigma a \right).
\]
(19)
Similar to the case of DBMT, we can exploit the block-triangular structure of the Hessian in (12) to carry out the E-step efficiently using forward filtering and backward smoothing.

2) M-step: Once the conditional expectation of the log-likelihood in (18) given \(\psi_{1:n}, R_0, \theta|0, \nu|0\) is available, we can update \(R^{[t+1]}\) and \(\theta|t+1\) using similar closed form equations.
as in [15]. But updating \( \nu^{[l+1]} \) by maximizing the conditional expectation of the log likelihood in \([18]\) w.r.t. \( \nu^{[l+1]} \) requires solving following nonlinear equation:

\[
\nu^{[l+1]} = \frac{1 - \log 2 + \frac{1}{N} \sum_{n=1}^{N} \left( \psi_n - \Omega_n \right) - \log \left( \nu^{[l+1]} \right)}{\frac{1}{N} \sum_{n=1}^{N} 1^T_j \exp \left( \psi_n - \Omega_n \right)},
\]

where \( F (\cdot) \) is the digamma function. We can use Newton’s method to solve this equation up to a given precision. An implementation of the log-DBMT is given by Algorithm 2.

**Algorithm 2** The log-DBMT Estimate of the \( k \)th log-Eigen-spectra

1. Initialize: observations \( \psi_1^{[k]} \), initial guess \( s_1^{[0]} \), initial guess \( R^{[0]} \), initial conditions \( 0 \) for \( t \in (0, 10^{-5}) \), Maximum Number of iteration \( L_{\text{max}} \).
2. repeat
3. \( l = 0 \).
4. Forward filter for \( n = 1, 2, \ldots, N \):
   \[ s_{n-1} = \theta[n] s_{n-2} \]
   \[ \Omega_{n-1} = \theta[n] \Omega_{n-2} + R^{[n]} \]
5. Backward smoother for \( n = N - 1, N - 2, \ldots, 1 \):
   \[ A_n = \theta[n] A_{n+1} \]
   \[ s_n = s_{n+1} + \Omega_n s_{n+1} \]
6. Let \( \bar{S}^{[l]} = [s_1^{[l]} \cdots s_N^{[l]}]^T \).
7. Update \( \nu^{[l+1]} \) and \( R^{[l+1]} \) as:
   \[
   \nu^{[l+1]} = \frac{1 - \log 2 + \frac{1}{N} \sum_{n=1}^{N} \left( \psi_n - \Omega_n \right) - \log \left( \nu^{[l+1]} \right)}{\frac{1}{N} \sum_{n=1}^{N} 1^T_j \exp \left( \psi_n - \Omega_n \right)},
   \]
   \[
   R^{[l+1]} = \frac{1}{N} \sum_{n=1}^{N} [s_n, R_{s,s}^{[l]} + \Omega_n + \Omega_n - \Omega_n] \]
8. Set \( l \leftarrow l + 1 \).
9. until \( \| S^{[l]} - S^{[l-1]} \|_2 < \text{tol} \) or \( l = L_{\text{max}} \).
10. Output: Denoised log-eigen-spectra \( \bar{S}^{[l]} \) where \( L \) is the index of the last iteration of the algorithm, and covariance matrices \( \Omega_n \) for \( n = 1, 2, \ldots, N \) in from last iteration of the algorithm.

Note that unlike the DBMT algorithm which pertains to a Gaussian observation model, the forward filtering step to compute \( s_{n-1} \) is nonlinear, and standard techniques such as Newton’s method can be used to solve for \( s_{n-1} \). We use the log-DBMT algorithm to find all the \( K \) log-Eigen-spectra and construct the log-DBMT estimate as:

\[
\hat{D}_n(f_j) = \frac{1}{K} \sum_{k=1}^{K} \exp \left( \frac{1}{N} \sum_{n=1}^{N} \left( \psi_n^{(k)} - \Omega_n \right) \right),
\]

where \( f_j := \frac{2\pi(j-1)}{N} \) for \( j = 1, 2, \ldots, J \) and \( n = 1, 2, \ldots, N \). Again, confidence intervals can be computed by mapping the Gaussian confidence intervals for \( s_n^{(k)} \)’s to the final log-DBMT estimate.

**C. Parameter Selection**

The window length \( W \), design bandwidth \( B \), and the number of tapers \( K \) need to be carefully chosen. Since both proposed algorithms are motivated by the classical MTM, we use the same guidelines for choosing these parameters [8]. The window length \( W \) is determined based on the expected rate of change of the PSD (given domain-specific knowledge) in order to make sure that the process remains approximately second-order stationary within each window. The design bandwidth \( B \) is chosen small enough to be able to resolve the dominant frequency components in the data, while being large enough to keep the time-bandwidth product \( \rho := \frac{W B}{6} \geq 1 \). The number of tapers \( K \) is then chosen as \( K \leq 2\rho - 1 \).

**IV. APPLICATION TO SYNTHETIC AND REAL DATA**

Before presenting our theoretical analysis, we examine the performance of DBMT and log-DBMT spectrogram estimators on synthetic data, and then demonstrate their utility in two real world data applications, namely spectral analysis of human EEG during sleep and Electric Network Frequency signal detection.

**A. Application to Synthetic Data**

The synthesis data consists of the linear combination of two amplitude-modulated and frequency-modulated processes with high dynamic range (i.e., high-Q). The amplitude-modulated component \( y_1^{(1)} \) is generated through modulating an AR(6) process tuned around 11 Hz by a cosine at a low frequency \( f_0 = 0.02 \) Hz. The frequency-modulated component \( y_2^{(2)} \) is a realization of an ARMA(6, 4) with varying pole location. To this end, the process has a pair of 3rd order poles at \( \omega_i := \frac{2\pi f_i}{t} \) and \( -\omega_i \), where \( f_i \) increases from 5 Hz, starting at \( t = 0 \), every \( \approx 26 \) s by increments of 0.48 Hz, to achieve frequency modulation. In summary, the noisy observations are given by:

\[
y_t = y_1^{(1)} \cos(2\pi f_0 t) + y_2^{(2)} + \sigma v_t,
\]

where \( v_t \) is a white Gaussian noise process and \( \sigma \) is chosen to achieve an SNR of 30 dB. The process is truncated at 600 s to be used for spectrogram analysis. Figure 2 shows a 12 second sample window of the process. Figure 3 shows the true as well as estimated spectrograms by the sliding window MTM, DBMT and log-DBMT estimators. Each row consists of three

![Fig. 2: Sample from the synthetic data from t = 244 s to t = 256 s.](image-url)
panels: the left panel shows the entire spectrogram; the middle panel shows a zoomed-in spectrotemporal region marked by the dashed box in the left panel; and the right panel shows the PSD at a selected time point marked by a dashed vertical line in the middle panel. Figure 3A shows true spectrogram of the synthetic process, in which the existence of both amplitude and frequency modulations make the spectrogram estimation a challenging problem.

Fig. 3B shows the MTM spectrogram estimate. We used windows of length 6 s and the first 3 tapers corresponding to a time-bandwidth product of 3 and 50% overlap to compute the estimates (note that the same window length, tapers and time-bandwidth product are used for the DBMT and log-DBMT estimators). Although the multitaper spectrogram captures the dynamic evolution of both components, it is blurred by the background noise and picks up spectral artifacts (i.e., vertical lines) due to window overlap, frequency mixing, and sampling noise. Fig. 3C demonstrates how the DBMT spectrogram estimate overcomes these deficiencies of the multitaper spectrogram: the spectrotemporal localization is sharper and smoother across time, no artifacts are picked due to overlapping windows, and frequency mixing is further mitigated. By comparing the second and third rows, two important observations can be made: first, as is revealed in the right panel, the DBMT captures the true dynamic range of the original noiseless PSD, while the MTM estimate fails to do so. Second, the confidence intervals (gray hulls) in Fig. 3C as compared to 3B are wider when the signal is weak (e.g., near 5 Hz) and tighter when the signal is strong (e.g., near 11 Hz. The latter observation highlights the importance of the model-based confidence intervals in interpreting the denoised estimates of DBMT: while the most likely estimate (i.e., the mean) captures the true dynamic range of the noiseless PSD, the estimator does not preclude cases in which the noise floor of −40 dB is part of the true signal, while showing high confidence in detecting the spectral content of the true signal that abides by the modeled dynamics.

Finally, Fig. 3D shows the log-DBMT spectrogram estimate, which shares the artifact rejection feature of the DBMT spectrogram. However, the log-DBMT estimate is smoother than both the multitaper and DBMT spectrograms in time as well as in frequency (see the zoomed-in middle panels), due to its sampling noise mitigation feature (by design). In addition, similar to the multitaper estimate, the log-DBMT estimator is not able to denoise the estimate by removing the observation noise. Though, the confidence intervals of the log-DBMT PSD estimate are tighter than those of the MTM estimate due to averaging across multiple windows via Bayesian filtering/smoothing. As we will show in Section V these qualitative observations can be established by our theoretical analysis. In the spirit of easing reproducibility, we have deposited a MATLAB implementation of these algorithms on the open source repository GitHub [50], which generates Figure 3.

B. Application to EEG data

To illustrate the utility of our proposed spectrogram estimators, we apply them to human EEG data during sleep. The data set is available online as part of the SHHS Polysomnography Database (https://www.physionet.org/ptdb/shpsgb/). The EEG data is 900 s long during stage 2 sleep, and sampled at 250 Hz. During stage 2 sleep, the EEG is known to manifest delta waves (0–4 Hz) and sleep spindles (transient wave packets with frequency 12–14 Hz) [31, 32]. Accurate localization of these spectrotemporal features has significant applications in studying sleep disorders and cognitive function [33]. Since the transient spindles occur at a time...
scale of seconds, we choose a window length of 2.25 s for all algorithms (with 50\% overlap for the MTM estimate). We also chose a time-bandwidth product of 2.25 for all algorithms, in order to keep the frequency resolution at 2 Hz. Figs. A, B and C show the multitaper, DBMT and log-DBMT spectrogram estimates, respectively, with a similar presentational structure as in Fig. 4. As the middle panels reveal, the MTM estimate is not able to clearly distinguish the delta waves and sleep spindles due to high background noise. The DBMT estimate shown in Fig. 4B, however, provides a significantly denoised spectrogram, in which the delta waves and sleep spindles are visually separable. The log-DBMT estimator shown in Fig. 4C provides significant spectrotemporal smoothing, and despite not fully reducing the background noise, provides a clear separation of the delta waves and spindles (see the PSD in the right panel). Similar to the analysis of synthetic data, the same observations regarding the confidence intervals of the estimators can be made.

C. Application to ENF data

Finally, we examine the performance of our proposed algorithms in tracking the Electrical Network Frequency (ENF) signals from audio recordings. The ENF signal corresponds to the supply frequency of the power distribution network which is embedded in audio recordings [34], [35]. The instantaneous values of this time-varying frequency and its harmonics form the ENF signal. The ability to detect and track the spectrotemporal dynamics of ENF signals embedded in audio recordings has shown to be crucial in data forensics applications [34].

Fig. 5A shows the spectrogram estimates around the sixth harmonic of the nominal 60 Hz ENF signal (data from [35]). We used 1000 s of audio recordings, with windows of length 5 s and using the first 3 tapers corresponding to a time-bandwidth product of 3 (with 25\% overlap for the MTM estimate). The two dominant components around the sixth ENF harmonic exhibit temporal dynamics, but are hard to distinguish from the noisy background. Fig. 5B shows the DBMT

---

**Fig. 4:** Spectrogram analysis of the EEG data. (A) MTM estimates, (B) DBMT estimates, and (C) log-DBMT estimates. Left: spectrograms. Middle: zoomed-in views from t = 700 s to t = 750 s. The color scale is in decibels. Right: PSD estimate corresponding to a window of length 2.25 s starting at t = 722.25 s. Grey hulls show 95\% confidence intervals.

**Fig. 5:** Spectrogram analysis of the ENF data. (A) MTM estimates, (B) DBMT estimates, and (C) log-DBMT estimates. Left: spectrograms. Middle: zoomed-in views from t = 480 s to t = 540 s. The color scale is in decibels. Right: PSD estimate corresponding to a window of length 5 s starting at t = 505 s. Grey hulls show 95\% confidence intervals.
spectrogram, in which the background noise is significantly suppressed, yielding a crisp and temporally smooth estimate of the ENF dynamics. The log-DBMT estimate is shown in Fig. 5; which provides higher spectrotemporal smoothness than the MTM. Although the log-DBMT shows smaller variability in the estimates (middle and right panels), the gain is not as striking as in the cases of synthetic data and EEG analysis, due to the usage of longer windows which mitigates the sampling noise for all algorithms. Similar observations as in the previous two cases regarding the statistical confidence intervals can be made, which highlight the advantage of modeling the spectrotemporal dynamics in spectrogram estimation.

V. THEORETICAL ANALYSIS

A. Filter Bank Interpretation

In order to characterize the spectral properties of any non-parametric spectrum estimator, the tapers applied to the data need to be carefully inspected. In the multitaper framework, the dpps sequences are used as tapers, which are known to produce negligible side-lobes in the frequency domain [7], [18]. The DBMT and log-DBMT algorithms also use the dpps tapers to alleviate the problem of frequency leakage. However, the estimate at each window is a function of all the window. Therefore, the theoretical properties of the multitaper method do not readily apply to our estimators.

To characterize the statistical properties of our estimates, we first need take a detour from the usual analysis of spectrum estimation techniques. In what follows, we mainly focus on the DBMT algorithm for the sake of presentation. First, we assume with no loss of generality that the window length \( W \) is an integer multiple of \( J \), the number of discrete frequencies, i.e., \( W = rJ \), for some \( r \in \mathbb{N} \), so that \( F_n = F_1, \forall n \). Second, we assume that the state noise covariance matrices are time-invariant, i.e., \( Q_n = Q, \forall n \). By virtue of the FIS procedure, and in the absence of any taper, it can be shown that the DBMT algorithm maps the entire data \( y := [y_1, y_2, \ldots, y_T]^T \) to the vector of coefficients \( \hat{X} := [\hat{X}_1^H, \hat{X}_2^H, \ldots, \hat{X}_N^H]^T \) according to [14]:

\[
\hat{X} = \lim_{l \to \infty} \hat{X}^{[l]} = GF^H y, \tag{23}
\]

where \( F \) is a block-diagonal matrix with \( F_1 \) as the diagonal blocks and \( G \) is a weighting matrix which depends only on \( Q_\infty = \lim_{l \to \infty} Q^{[l]}, \alpha_\infty = \lim_{l \to \infty} \alpha^{[l]} \), and window length, \( W \). The rows of \( GF^H \) forms a filter bank whose output is equivalent to the time-frequency representation. The use of dpps tapers modifies the filter banks as:

\[
\hat{X}^{(k)} = G^{(k)} F^H U^{(k)} y, \tag{24}
\]

where \( U^{(k)} \) is another block diagonal matrix with \( U^{(k)} := \text{diag}[u^{(k)}] \) on its diagonal blocks. To see the structure of the weighting matrix \( G^{(k)} \), by expanding \( x^{(k)}_{n|N} \) in terms of data [14], one obtains:

\[
x^{(k)}_{n|N} = \sum_{s=1}^{n-1} \prod_{m=s}^{n-1} [\alpha (I-K_mF_m)]K_sU^{(k)}y_s + K_nU^{(k)}y_n + \sum_{s=n+1}^{N} \prod_{m=n}^{s-1} B_mK_sU^{(k)}y_s \tag{25}
\]

In order to continue our analysis, we make two assumptions common in the analysis of adaptive filters [16], [17]. First, we assume that the parameter estimates \( Q_\infty \) and \( \alpha_\infty \) are close enough to the true values of \( Q \) and \( \alpha \), and therefore replace them by \( Q \) and \( \alpha \), i.e., as if the true parameters were known. Note that we have discarded the dependence of \( Q \) and \( \alpha \) on \( k \) for lucidity of analysis. Second, noting that \( \alpha (I-K_mF_m) = \alpha \Sigma_m |m \Sigma_m^{-1}| \) and \( B_m = \alpha \Sigma_m |m \Sigma_m^{-1}| \) and that in steady state we have \( \Sigma_m |m \Sigma_m^{-1} = \alpha^2 \Sigma_\infty + Q \), Eq. (25) can be approximated by:

\[
x^{(k)}_{n|N} = \sum_{s=1}^{N} \Lambda^{s-n} |I \Gamma F^H U^{(k)}y_s, \tag{26}
\]

for \( 1 \ll n \ll N \), where \( \Lambda = \alpha \Sigma_\infty (\alpha^2 \Sigma_\infty + Q)^{-1} \), and \( \Gamma = (\alpha^2 \Sigma_\infty + Q)(I - rW ((\alpha^2 \Sigma_\infty + Q)^{-1} + rW I)^{-1}) \). That is, for values of \( n \) far from the data boundaries, the weighting matrix is equivalent to a weighted set of dpps tapers in matrix form acting on all the data windows with an exponential decay with respect to the \( n \)th window.

As an example, the equivalents filters of the DBMT estimator for the 11Hz and 9Hz frequencies around 300 secs, from the synthetic data example are shown in Fig. 6. They are also compared to the equivalent filters corresponding to MTM in the frequency domain. As apparent from Fig. [6] the weighting matrix sets the gain of these filters in an adaptive fashion across all windows, unlike the MTM which only uses the data in window \( n \). In addition, the filter corresponding to

![Fig. 6: Equivalent filters corresponding to the DBMT estimate of the synthetic data example. Left: equivalent filters in time around \( t = 300 \) s. Right: equivalent filters of DBMT (red) and MTM (green) in frequency.](image)
frequency of 9Hz, which is negligible in the data, is highly attenuated, resulting in significant noise suppression. In this sense, the proposed estimation method can be viewed as a data-driven denoising method for constructing time-frequency representations given noisy time series data. Next, we will characterize the performance of the DBMT estimator in terms of bias-variance trade-off.

B. Bias and Variance Analysis

We first consider the implication of the stochastic continuity constraint of Eq. (5) on the evolution of the orthogonal increment processes governing the time series data. We first assume that the parameters \( \alpha(k) = \alpha \), for all \( k = 1, 2, \cdots, K \). Suppose that the data in window \( n \) has a Cramér representation with an orthogonal increment process \( dz_n(f) \), \( n = 1, 2, \cdots, N \). Then, one way to achieve the stochastic continuity of Eq. (5) is to assume:

\[
dz_{n+1}(f) = \alpha dz_n(f) + de_n(f),
\]

where \( de_n(f) \) is a Gaussian orthogonal increment process, independent of \( dz_n(f) \). In the forthcoming analysis we also assume the local stationarity condition, i.e., the generalized Fourier transform of the process remains constant within each window. This assumption is common in the analysis of nonparametric spectral estimators \cite{8, 17}. Finally, we assume a scaling of \( K, N, W \to \infty, B \to 0, BW \to \rho \), for some constant \( \rho \) \cite{38}. The following theorem characterizes the bias and variance of the DBMT estimator:

**Theorem 1.** Suppose that the locally stationary process \( y_t \) is governed by orthogonal increment processes evolving according to the dynamics \( dz_{n+1}(f) = \alpha dz_n(f) + de_n(f) \), with \( \alpha < 1 \), where the noise process \( de_n(f) \), \( \forall f \in \{-1/2, 1/2\} \) is a zero-mean Gaussian increment process with variance \( \sigma^2 \), independent of \( dz_n(f) \). If the process is corrupted by additive zero-mean Gaussian noise with variance \( \sigma^2 \), then for \( f \in \{f_1, f_2, \cdots, f_j\} \), the DBMT estimate satisfies:

\[
\mathbb{E}[\tilde{D}(f)] - D(f) \leq \left(1 - \frac{1}{K} \sum_{k=1}^{K} \lambda_k \right) \kappa_n(f) \sup_{f} \{D(f)\} + \left|1 - \kappa_n(f)\right| D(f) + \mu_n(f)(\alpha^2 + \kappa_n(f)\alpha(1)),
\]

where \( \lambda_k \) is the eigenvalue associated with the kth PSWF \( D(f) := q(f)/(1 - \alpha) \), and \( \kappa_n(f), \mu_n(f) \) are functions of \( \alpha \) and \( q(f) \) and explicitly given in the proof.

**Theorem 2.** Under the assumptions of Theorem 1, the variance of the DBMT estimate \( \tilde{D}(f) \) satisfies:

\[
\text{Var}\{\tilde{D}(f)\} \leq \frac{2}{K} \sup_{f} \{\kappa_n(f)D(f) + \mu_n(f)\alpha^2\}^2.
\]

The proofs of Theorems 1 and 2 integrate the treatment of \cite{38} with the structure of the FIS estimates, and are presented in Appendix A. In order to illustrate the implications of these theorems, several remarks are in order:

**Remark 1.** The function \( \kappa_n(f) \) controls the trade-off between bias and variance: for values of \( \kappa_n(f) < 1 \), the bound on the variance decreases while the bias bound increases, and for \( \kappa_n(f) \approx 1 \), all the terms in the bias bound become negligible, while the variance bound increases. The function \( \mu_n(f) \), on the other hand, reflects observation noise suppression in both the bias and variance. Note that these upper bounds are tight and achieved for a signal with flat spectrum.

**Remark 2.** The bias and variance bounds of \cite{38} for the MTM can be recovered by setting \( \kappa_n(f) = 1, \mu_n(f) = 1, \) and \( \sigma^2 = 0 \) in the results of Theorems 1 and 2, i.e., in the absence of signal dynamics and observation noise. For the DBMT estimator, the signal and observation noise variances, respectively contribute to the bias/variance upper bounds in different fashions through \( \kappa_n(f) \) and \( \mu_n(f) \), due to the distinction of the signal and observation noise in our state-space model. In contrast, in the classical MTM, possible observation noise is treated in the same way as the true data, and hence both the signal and noise variances have equal contributions in the estimator bias/variance.

The functions \( \kappa_n(f) \) and \( \mu_n(f) \) do not have closed-form expressions with respect to the state-space parameters \( \alpha, \sigma^2 \) and \( q(f) \). In order to illustrate the roles of \( \kappa_n(f) \) and \( \mu_n(f) \), we consider the scenario under which the upper bounds on the bias and variance are achieved, i.e., \( q(f) \) being independent of \( f \), and hence \( \kappa_n(f) = \kappa_n \) and \( \mu_n(f) = \mu_n, \forall f \). In this scenario, even though the dependence of \( \mu_n \) and \( \kappa_n \) on the state-space parameters is quite involved, it is possible to obtain upper and lower bounds on \( \kappa_n \) and \( \mu_n \). As it is shown in Proposition \cite{38} in Appendix B, the main parameters determining the behavior of \( \kappa_n \) and \( \mu_n \) are \( q/\sigma \) (i.e., the SNR) and \( \alpha \) (i.e., temporal signal dependence). Here, we present a numerical example for clarification. Fig. 7 shows the plot of \( \kappa_n \) vs. \( \alpha \) for \( n = 50 \) and \( q/\sigma^2 = 10 \). It is apparent that \( \mu_n \) increases with \( \alpha \) and does not exceed 1. The fact that \( \mu_n < 1 \) implies that the DBMT estimator achieves a higher noise suppression compared to the classical MTM. This fact agrees with the noise suppression performances observed in Section IV-A.

Fig. 7 shows the plot of \( \kappa_n \) vs. \( \alpha \), which exhibits a similar increasing trend, but eventually exceeds 1. This result implies that with a careful choice \( \alpha \), it is possible to achieve \( \kappa_n < 1 \), and hence obtain lower variance than that of the classical MTM estimator. Fig. 8 illustrates this statement by showing the value of \( \alpha \) for which \( \kappa_n \approx 1 \) vs. \( q/\sigma^2 \). For models with high temporal dependence (i.e., \( \alpha \) close to 1), it is possible to achieve \( \kappa_n < 1 \) and hence reduce the estimator variance, due to the increase in the weight of data pooled from adjacent windows, even for small values of \( q/\sigma^2 \) (i.e., low SNR). However, this reduction in variance comes with the cost of increasing the bias. On the contrary when the data across windows have low temporal dependence (i.e., \( \alpha \ll 1 \), it

![Fig. 7: (A) \( \mu_n \) versus \( \alpha \), (B) \( \kappa_n \) and its upper/lower bounds versus \( \alpha \) for \( N = 100, n = 50 \), and \( q/\sigma^2 = 10 \).](image-url)
is only possible to achieve a reduction in variance for high values of $q_{\alpha}/\sigma^2$ (i.e., high SNR). This is due to the fact that at low SNR with low temporal dependence, pooling data from adjacent windows is not beneficial in reducing the variance in a particular window, and indeed can result in higher bias.

Remark 3. Even though the parameter $\alpha$ is estimated in a data-driven fashion, it can be viewed as a tuning parameter controlling the bias-variance trade-off, given the foregoing discussion. For a given SNR, fixing $\alpha$ at a small value can help reduce the variance but with a cost of increasing bias, and vice versa. In light of this observation, Fig. 8 can be thought of as a guideline for choosing $\alpha$ to achieve $\kappa_{\alpha} \approx 1$, so that the estimator is nearly unbiased, and achieves a lower variance than that of the classical multitaper estimator due to the noise suppression virtue of the state-space model. Although we focused on the case of flat spectrum in the foregoing discussion, it is possible to numerically compute these trade-off curves for more general cases, given the general expressions for $\kappa_{\alpha}(f)$ and $\mu_{\alpha}(f)$ given in Appendix A.

Remark 4. Extending Theorems 1 and 2 to the log-DBMT algorithm is not straightforward, due to the high nonlinearity of the underlying state-space models. However, under the common Gaussian approximation of the log-posterior density, the well-known variance reduction property of the fixed interval smoother [25, 37] carries over to the estimator of log $S_{\alpha}(f)$. That is, the variance of the log-DBMT estimate of log $S_{\alpha}(f)$ obtained using the state-space model is lower than that of the classical MTM estimate which only uses the data within window $n$. This fact agrees with our earlier observation in Section V-A regarding the tightening of the confidence intervals for log-DBMT as compared to the MTM estimates.

VI. CONCLUDING REMARKS

Spectral analysis of nonstationary time series data poses serious challenges for classical nonparametric techniques, in which temporal smoothness of the spectral representations are implicitly captured using sliding windows with overlap. This widely-practiced approach tends to ignore the inherent smoothness of the data and is not robust against measurement/sampling noise. In this paper, we address these issues and provide an alternative to the sliding window spectrogram analysis paradigm. We propose two nonparametric spectrogram estimators, namely the DBMT and log-DBMT estimators, by integrating techniques from multitaper analysis and Bayesian estimation. To this end, we explicitly model the temporal dynamics of the spectrum using a state-space model over the spectral features obtained by multitapering. Therefore our algorithms inherit the optimality features of both Bayesian estimators and multitaper analysis.

Our algorithms admit efficient and simple implementations, thanks to the Expectation-Maximization algorithm and the well-known fixed interval state-space smoothing procedure. Unlike existing approaches, our algorithms require no a priori assumptions about the structure of the spectral representation and operate in a fully data-driven fashion. While both algorithms yield spectral representations that are continuous in time, by design the DBMT algorithm significantly suppresses the measurement noise in forming the spectrogram and the log-DBMT algorithm mitigates sampling noise due to small observation length. We establish the performance gains provided by our algorithms through theoretical analysis of the bias-variance trade-off, as well as application to synthetic and real data from human EEG data during sleep and ENF signals from audio recordings.

APPENDIX A

PROOFS OF THEOREMS 1 AND 2

Recall from Eq. (26) that the $k$th eigen-coefficient estimate at window $n$ can be written in terms of the observed data as:

$$\hat{x}_{n}^{(k)} = \sum_{s=1}^{\infty} \Lambda^{s-n} \Gamma F_{s}^{\dagger} U_{s}^{(k)},$$

(28)

where $y_{s}^{(k)} = U_{s}^{(k)} \odot y_{s} = U_{s}^{(k)} y_{s}$. Given that $Q$ is a diagonal matrix with elements $q(j_{f}), j = 1, 2, \ldots, J$, it can be shown that $\Sigma_{\infty}$, $\Lambda$ and $\Gamma$ are also diagonal matrices. Denoting the elements of $\Sigma_{\infty}$, $\Gamma$ and $\Lambda$, respectively by $\tau_{j}(f)$, $\gamma_{j}(f)$, and $\alpha_{j}(f)$ for $f \in \{f_{1}, f_{2}, \ldots, f_{J}\}$, we have

$$\eta(f) = \frac{\alpha^{2} \tau_{j}(f) + q(f)}{1 + \tau_{j}(f) + q(f)}.$$  

A. Proof of Theorem 1

First note that Eq. (27) implies that $E[d_{\alpha}(f) d_{\alpha+1}(f')] = \alpha^{2} D_{\alpha}(f - f') ddf'$. By invoking the Cramer representation, the covariance of the data tapered by the $k$th and $l$th dpps sequences can be expressed as:

$$E \left[ (y_{s}^{(k)}) (y_{s}^{(l)})^{\dagger} \right] = \alpha^{2} s^{-l} \int_{-1/2}^{1/2} U_{k}(f_{j} - \beta) D(\beta) U_{l}^{\dagger}(f_{j} - \beta) d\beta + \int_{-1/2}^{1/2} U_{k}(f_{j} - \beta) \sigma^{2} s^{-l} U_{l}^{\dagger}(f_{j} - \beta) d\beta.$$  

(29)

From Eqs. (16) and (28), we get:

$$\hat{D}_{n}(f) = \frac{\eta^{2}(f)}{K} \sum_{k=1}^{K} \sum_{s=1}^{N} \gamma(f_{s}^{|s-n|})(f_{s})^{|s-n|}(y_{s}^{(k)}) (y_{s}^{(k)})^{\dagger}.$$  

(30)

Taking the expectation of both sides and after some simplification, one arrives at:

$$E \left[ \hat{D}_{n}(f) \right] = \frac{\eta^{2}(f)}{K} \sum_{k=1}^{K} \sum_{s=1}^{N} \gamma(f_{s}^{|s-n|})(f_{s})^{|s-n|}E \left[ (y_{s}^{(k)}) (y_{s}^{(k)})^{\dagger} \right] = \left[ \eta(f_{j})^{2} \sum_{s=1}^{N} \gamma(f_{s}^{|s-n|})(f_{s})^{|s-n|} \alpha^{2} s^{-l} \right] \times \frac{1}{K} \sum_{k=1}^{K} \int_{-1/2}^{1/2} U_{k}(f_{j} - \beta) D(\beta) U_{k}^{\dagger}(f_{j} - \beta) d\beta + \left[ \eta(f_{j})^{2} \sum_{s=1}^{N} \gamma(f_{s}^{|s-n|}) \right] \sigma^{2}.$$  

(31)
Using the orthogonality of the PSWFs as in [38], and using the fact that \(D(\beta) \leq \sup_f D(f), \forall \beta\), we get:

\[
|\mathbb{E}[\hat{D}_n|N(f)] - \kappa_n(f)D(f)| \leq \kappa_n(f)(\sup \{D(f) - D(f)\} \times \\
(1 - \frac{1}{K} \sum_{k=1}^{K} \lambda_k) + \mu_n(f)\sigma^2 + \kappa_n(f)\sigma(1),
\]

(32)

where

\[
\kappa_n(f) := \eta(f)^2 \sum_{s=1}^{N} \sum_{s' \neq 1}^{N} (\gamma(f)|s-n|\gamma(f)|s'-n|\alpha|s-s'|),
\]

\[
\mu_n(f) := \eta(f)^2 \sum_{s=1}^{N} \gamma(f)^2|s-n|,
\]

for \(f \in \{f_1, f_2, \ldots, f_J\}\). Using the triangle inequality, the bound of Theorem [41] on \(|\mathbb{E}[\hat{D}_n|N(f)] - D(f)\|\) follows. ■

B. Proof of Theorem 2

Using the notation of Appendix A, we have:

\[
\text{Cov}\left\{\hat{D}_n^{(k)}(f), \hat{D}_m^{(l)}(f')\right\} = \\
\eta(f)^4 \sum_{s,s',t',t}^{N} \gamma(f)^{s-n}\gamma(f)^{s'-n}\gamma(f')^{t-m}\gamma(f')^{t'-m}\times \\
\left[\alpha^{s-t}\alpha^{s'-t'}\int_{D(\beta)} U_k(f - \beta)U_l(f' + \beta)U_k(-\beta' - f)\times \\
U_l(\beta' - f')(D(\beta) + \sigma^2\delta(s-t))(D(\beta') + \sigma^2\delta(s'-t'))d\beta d\beta' + \\
\alpha^{s-t}\alpha^{s'-t'}\int_{D(\beta)} U_k(f - \beta)U_l(f' - \beta)U_k(-\beta' - f)\times \\
U_l(\beta' + f')(D(\beta) + \sigma^2\delta(s'-t'))d\beta d\beta'.
\]

(33)

Note that we have omitted the integral limits, as they are understood to be same as in [41] henceforth. After summing over all tapers and rearranging the summations, the first expression within the brackets in [33] becomes:

\[
\int \left[\eta(f)^2 \sum_{s,t}^{N} \gamma(f)^{s-n}\gamma(f')^{t-m}\alpha|s-t|\right] (D(\beta) + \sigma^2\delta(s-t)) \times \\
\left[\eta(f)^2 \sum_{s',t'}^{N} \gamma(f)^{s'-n}\gamma(f')^{t'-m}\alpha|s'-t'|\right] (D(\beta') + \sigma^2\delta(s'-t')) \times \\
\sum_{k=1}^{K} U_k(f - \beta)U_k(-\beta' - f) \sum_{l=1}^{K} U_l(f' + \beta)U_l(\beta' - f)d\beta d\beta'.
\]

(34)

Let

\[
\mathcal{A}(n, m, f) := \left[\eta(f)^2 \sum_{s,t}^{N} \gamma(f)^{s-n}\gamma(f')^{t-m}\alpha|s-t|\right] D(f) \\
+ \left[\eta(f)^2 \sum_{s=1}^{N} \gamma(f)^2|s-n|\right] \sigma^2
\]

(35)

Using the Schwarz inequality, the integral in Eq. (34) can be bounded by:

\[
\int \left[\int \sum_{k=1}^{K} U_k(f - \beta)U_k(-\beta' - f) \mathcal{A}(n, m, \beta)\mathcal{A}(n, m, \beta')d\beta d\beta' \times \\
\int \left[\sum_{l=1}^{K} U_l(f' + \beta)U_l(\beta' - f') \mathcal{A}(n, m, \beta)\mathcal{A}(n, m, \beta')d\beta d\beta' \right]^{1/2}.
\]

(36)

Using bounds on the convolutions of PSWFs form [38], and upper bounding \(\mathcal{A}(n, m, \beta)\) by \(\sup_f \{\kappa_n(f)D(f) + \mu_n(f)\sigma^2\}\), the statement of the theorem on the variance of the DBMT estimate \(\hat{D}_n(f)\) follows.

Appendix B

Characterization of \(\kappa_n(f)\): Bounds and Parameter Dependence

A. Lower and Upper Bounds on \(\kappa_n(f)\)

Consider the scenario where \(q(f) = q\), i.e., flat spectrum. Then, the dependent of \(\gamma(f)\) and \(\kappa_n(f)\) on \(f\) is suppressed. We have the following bound on \(\kappa_n:\)

**Proposition 3.** For \(0 < \gamma, \alpha < 1\), the quantity \(\kappa_n\) can be bounded as:

\[
\kappa_n - \left(1 - \frac{\gamma}{\alpha}\right)^2 \left[\frac{1 + \alpha^2 - 2(\alpha\gamma)^N}{1 - \alpha^2}\right] + \frac{\gamma}{(1 - \gamma)^2} \leq \left(1 - \frac{\gamma}{\alpha}\right)^2 \frac{\gamma}{(1 - \gamma)^2}.
\]

where \(T_0 := \frac{1 + \gamma^2 - \gamma^N}{1 - \gamma^2}\).

**Proof:** To get an upper bound on \(\kappa_n\), we rewrite the expression defining \(\kappa_n\) as:

\[
\kappa_n = \eta^2 \sum_{s=1}^{N} \sum_{t=1}^{N} \gamma(s-n)\gamma(s'-n)\alpha|s-t|.
\]

Now, let us define \(T_t := \sum_{s=1}^{N} \gamma(s-n)\gamma(s'-n)\). Then it can be verified that:

\[
T_{t+1} = \gamma T_t, \text{ when } t \geq N - n \\
\leq \gamma T_t + \gamma^{t+1}, \text{ when } 0 \leq t < N - n
\]

(37)

and

\[
T_{t-1} = \gamma T_t, \text{ when } t \leq -n + 1 \\
\leq \gamma T_t + \gamma^{t-1}, \text{ when } -n + 1 < t < 0.
\]

(38)

Also, we have:

\[
\sum_{t=0}^{N-1} \alpha^t T_t \leq \sum_{t=0}^{N-1} \alpha^t T_0 + \sum_{t=1}^{N-n} \gamma^t + \sum_{t=N-n}^{N-1} (N-n)\gamma^t \\
\leq \frac{1 - (\alpha\gamma)^N}{1 - \alpha\gamma} T_0 + \frac{\gamma}{(1 - \gamma)^2}.
\]

(39)

Similarly, we have:

\[
\sum_{t=-N+1}^{0} \alpha^t T_t \leq \frac{1 - (\alpha\gamma)^N}{1 - \alpha\gamma} T_0 + \frac{\gamma}{(1 - \gamma)^2},
\]

(40)

which along with (39) leads to the claimed upper bound. For the lower bound, we use the fact that

\[
\gamma T_t = \begin{cases} 
T_{t+1} & \text{when } t \geq 0 \\
T_{t-1} & \text{when } t \leq 0
\end{cases}
\]

(41)

which implies \(\sum_{t=0}^{N-1} \alpha^t T_t \geq \frac{1 - (\alpha\gamma)^N}{1 - \alpha\gamma} T_0\) and

\[
\sum_{t=-N+1}^{0} \alpha^t T_t \geq \frac{1 - (\alpha\gamma)^N}{1 - \alpha\gamma} T_0.
\]

Using the latter lower bounds for \(\kappa_n\) yield the claimed lower bound. ■
B. Relation between $\kappa_n$ and $Q$

The expressions for $\kappa_n(f)$ and $\mu_n(f)$ in Appendix A depend on $\gamma(f)$ and $\alpha$. But $\gamma(f)$ itself depends on $q(f)$ and $\alpha$, and it is not straightforward to give a closed-form expression of $\gamma(f)$ merely in terms of $q(f)$ and $\alpha$, since it requires computation of the filtered error covariance matrix $\Sigma_{n|n}$ given $Q$. Again, by invoking the steady-state approximation, and defining $\Sigma_{n|n-1} = \Sigma$, $\forall n = 1, 2, \cdots, N$, the matrix $\Sigma$ can be obtained by solving the following algebraic Riccati equation:

$$\Sigma = \alpha^2 \Sigma - \alpha^2 \Sigma F_n H_n (2\alpha^2 I + H_n \Sigma F_n H_n) \Sigma + Q$$

(42)

and thereby the steady-state error covariance matrix $\Sigma_{n|\infty} = \Sigma_{\infty}$ is given by $\Sigma_{\infty} = \frac{1}{\alpha^2} (\Sigma - Q)$.

Although this procedure can be carried out numerically, in general it is not possible to solve the Riccati equation to get a closed-form expression for arbitrary $Q$. In order to illustrate the explicit dependent of $\kappa_n(f)$ on the state-space parameters, we consider the case of flat spectrum where $Q = qI$. In this case, it can be shown that $\Sigma = qI$ for some $\sigma > 0$ and the matrix equation (42) reduces to a simpler scalar equation for $\zeta$ given by:

$$\zeta = \alpha^2 \frac{\zeta - q}{\alpha^2} \left(1 - \frac{rW(\zeta/\alpha^2)}{rW(\zeta/\sigma^2) + 1}\right) + q$$

(43)

Following simplification, a quadratic equation for $\zeta$ results, and since $\zeta > 0$, the positive solution for $\zeta$ is given by:

$$\zeta = \frac{1}{2rW} \left(-\left(1-\alpha^2 - rW \frac{q}{\sigma} \right)^2 + 4rW^2 \frac{q}{\sigma^2} \right)^{1/2} + 1$$

(44)

Then, $\gamma$ can be computed as $\gamma = \frac{1}{\alpha^2}(1 - \frac{q}{\gamma/\sigma^2})$. Using these expressions for $\gamma$ and $\alpha$, the functions $\kappa_n$ and $\mu_n$ can be computed. Fig. 7B shows the upper and lower bounds on $\kappa_n$ evaluated for $n = 50$ and different values of $\alpha$ for $q/\sigma^2 = 10$. The upper and lower bounds are simple functions of $\alpha$ and $q/\sigma^2$ and can be used to further inspect the performance trade-offs of the DBMT algorithm with respect to the state-space model parameters.

ACKNOWLEDGMENT

We would like to thank Adi Haji-Ahmad and Min Wu for providing us the ENF data from [35].

REFERENCES

[1] T. F. Quatieri, *Discrete-time speech signal processing: principles and practice*. Prentice Hall, 2008.

[2] J. S. Lim, “Two-dimensional signal and image processing.” Englewood Cliffs, NJ: Prentice Hall, 1990. p. 710.

[3] G. Buzsaki, *Rhythms of the Brain*. Oxford University Press, 2009.

[4] W. J. Emery and R. E. Thomson, *Data analysis methods in physical oceanography*. Elsevier Science, 2001.

[5] M. Ghil, M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi et al., “Advanced spectral methods for climatic time series,” *Reviews of Geophysics*, vol. 40, no. 1, pp. 1003, 2002.

[6] Ö. Yilmaz, *Seismic data analysis: processing, inversion, and interpretation of seismic data*. SEG Books, 2001, no. 10.

[7] D. J. Thomson, “Spectrum estimation and harmonic analysis,” *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept 1982.

[8] B. Babadi and E. N. Brown, “A review of multitaper spectral analysis,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1555–1564, May 2014.

[9] B. Efron, *The jackknife, the bootstrap and other resampling plans*. SIAM, 1982.

[10] N. E. Huang et al., “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. R. Soc. Lond. A.*, vol. 454, no. 1971, pp. 903–995, 1998.

[11] I. Daubechies, J. Lu, and H.-T. Wu, “Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool,” *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 243–261, 2011.

[12] I. Daubechies, Y. G. Wang, and H.-T. Wu, “Conceput: concentration of frequency and time via a multitapered synchrosqueezed transform,” *Phil. Trans. R. Soc. A.*, vol. 374, no. 2065, p. 20150193, 2016.

[13] J. Xiao and P. Flandrin, “Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement,” *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2851–2860, 2007.

[14] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, “Robust spectrotemporal decomposition by iteratively reweighted least squares,” *Proc. Natl. Acad. Sci.*, vol. 111, no. 50, pp. E5336–E5345, 2014.

[15] M. Loeve, *Probability Theory*. London: D. Van Nostrand Co., 1963.

[16] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[17] D. J. Thomson, “Multitaper analysis of nonstationary and nonlinear time series,” in *Nonlinear and nonstationary signal processing*. London, UK: Cambridge Univ. Press, 2000, pp. 317–394.

[18] B. Efron and A. T. Walden, *Spectral Analysis for Physical Applications*. Cambridge University Press, 1993.

[19] D. Slepian, “Prolate spheroidal wave functions, fourier analysis, and uncertainty-v: the discrete case,” *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1430, May 1978.

[20] O. Rosen, D. S. Stoffer, and S. Wood, “Local spectral analysis via a bayesian mixture of smoothing splines,” *J. Amer. Statist. Assoc.*, vol. 104, no. 485, pp. 249–262, 2009.

[21] R. L. Prentice, “A log gamma model and its maximum likelihood estimation,” *Biometrika*, vol. 61, no. 3, pp. 539–544, 1974.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *J. R. Stat. Soc. Ser. B Statistical Methodology*, pp. 1–38, 1977.

[23] A. C. Smith and E. N. Brown, “Estimating a state-space model from point process observations,” *Neural Comput.*, vol. 15, pp. 965–991, August 2003.

[24] R. H. Shumway and D. S. Stoffer, “An approach to time series smoothing and forecasting using the em algorithm,” *Science*, vol. 3, no. 4, pp. 653–664, July 1982.

[25] H. E. Rauch, C. T. Striebel, and T. F., “Maximum likelihood estimates of liner dynamic systems,” *AIAA Journal*, vol. 3, pp. 1445–1450, August 1965.

[26] P. D. Jong and M. J. Mackinnon, “Covariances for smoothed estimates in state space models,” *Biometrika*, vol. 75, no. 3, p. 601, 1988.

[27] K. Lange, *Optimization*. Springer, 2004.

[28] G. Kitagawa, “Non-gaussian state-space modeling of nonstationary time series,” *J. Amer. Statist. Assoc.*, vol. 82, no. 400, pp. 1032–1041, 1987.

[29] L. Fahrmeir, “Posterior mode estimation by extended kalman filtering for multivariate dynamic generalized linear models,” *J. Amer. Statist. Assoc.*, vol. 87, no. 418, pp. 501–509, 1992.

[30] *Dynamic Bayesian Multitaper Spectral Estimators*. Available on GitHub Repository: https://github.com/proloyd/DBMT, 2017.

[31] L. D. Gennaro and M. Ferrara, “Sleep spindles: an overview,” *Sleep Medicine reviews*, vol. 7, no. 9, pp. 423–440, 2003.

[32] D. A. McCormick and T. J. Sejnowski, “Thalamocortical oscillations in sleep,” *Science*, vol. 262, no. 5134, pp. 679–85, October 1993.

[33] L. De Gennaro and M. Ferrara, “Sleep spindles: an overview,” *Sleep Medicine reviews*, vol. 7, no. 9, pp. 885–888, Sept 2013.

[34] S. Haykin, *Adaptive filter theory*, ser. Prentice-Hall information and system sciences series. Prentice Hall, 1991.

[35] B. D. Anderson and J. B. Moore, “Optimal filtering,” Englewood Cliffs, vol. 21, pp. 22–95, 1979.

[36] K. Li and M. Rosenblatt, “Prolate spheroidal spectral estimates,” *Statistics & Probability Letters*, vol. 78, no. 11, pp. 1339–1348, 2008.