Fast Multitaper Spectral Estimation

Santhosh Karnik, Justin Romberg, and Mark A. Davenport
Georgia Institute of Technology, School of Electrical and Computer Engineering
skarnik1337@gatech.edu, jrom@ece.gatech.edu, mdav@gatech.edu

***Abstract*—Thomson's multitaper method using discrete prolate spheroidal sequences (DPSSs) is a widely used technique for spectral estimation. For a signal of length N, Thomson's method requires selecting a bandwidth parameter W, and then uses $K \approx 2NW$ tapers. The computational cost of evaluating the multitaper estimate at N grid frequencies is $O(KN \log N)$. It has been shown that the choice of W and K which minimizes the MSE of the multitaper estimate is $W = O(N^{-1/5})$ and $K = O(N^{4/5})$. This choice would require a computational cost of $O(N^{9/5} \log N)$. We demonstrate an ϵ-approximation to the multitaper estimate which can be evaluated at N grid frequencies using $O(N \log^2 N \log \frac{1}{\epsilon})$ operations.**

I. Introduction

Let $x(t), t \in \mathbb{R}$ be a stationary, ergodic, zero-mean, Gaussian stochastic process. The Cramer representation of $x(t)$ is given by

$$x(t) = \int_{-1/2}^{1/2} e^{j2\pi ft}\, dZ(f),$$

and the spectral density of $x(t)$ is given by

$$S(f)\, df = \mathbb{E}\left[|\, dZ(f)|^2\right].$$

The problem of spectral estimation is to estimate $S(f)$ from N equally spaced samples

$$\boldsymbol{x} = \begin{bmatrix} x(0) & x(1) & \cdots & x(N-1) \end{bmatrix}^T \in \mathbb{C}^N.$$

Thomson's multitaper method for spectral estimation [6] can be described as follows. For a given half-bandwidth parameter $W \in (0, \frac{1}{2})$, we define the Slepian basis vectors $\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{N-1} \in \mathbb{R}^N$ as the orthonormal eigenvectors of the $N \times N$ prolate matrix $\boldsymbol{B}$, whose entries are given by[1]

$$\boldsymbol{B}[m,n] = \frac{\sin[2\pi W(m-n)]}{\pi(m-n)} \text{ for } m, n \in [N].$$

The eigenvectors are ordered such that corresponding eigenvalues $\lambda_0 > \lambda_1 > \cdots > \lambda_{N-1}$ are sorted in descending order. For each $k \in [N]$, we can use $\boldsymbol{s}_k$ as a taper to define a single tapered spectral estimate $\widehat{S}_k(f)$, i.e.,

$$\widehat{S}_k(f) = \left| \sum_{n=0}^{N-1} \boldsymbol{s}_k[n] \boldsymbol{x}[n] e^{-j2\pi fn} \right|^2.$$

Then, we pick an integer K and define the unweighted multitaper spectral estimate of $\boldsymbol{x}$ as

$$\widehat{S}_K^{\text{mt}}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \widehat{S}_k(f).$$

[1]For any integer N, we define $[N] := \{n \in \mathbb{Z} : 0 \le n < N-1\}$.

Since the first slightly less than $2NW$ Slepian basis vectors have spectra concentrated in $[-W, W]$, the number of tapers K is usually chosen to be slightly less than $2NW$. Thompson also considered the eigenvalue weighted multitaper spectral estimate [6]

$$\widehat{S}_K^{\text{eig}}(f) = \frac{\sum_{k=0}^{K-1} \lambda_k \widehat{S}_k(f)}{\sum_{k=0}^{K-1} \lambda_k}.$$

In many applications, it is desirable to estimate the spectrum on a grid of N evenly spaced frequencies, i.e., $f = \frac{m}{N}$ for $m \in [N]$. For each $k \in [K]$, evaluating $\widehat{S}_k(f)$ at all N grid frequencies takes $O(N \log N)$ operations via a length-N FFT of the elementwise product $\boldsymbol{s}_k \circ \boldsymbol{x}$. After this, only $O(KN)$ more operations are needed to evaluate the weighted/unweighted sum at all N grid frequencies. Hence, the total computation required to evaluate either $\widehat{S}_K^{\text{mt}}(f)$ or $\widehat{S}_K^{\text{eig}}(f)$ at the N grid frequencies can be done in $O(KN \log N)$ operations. Also, the cost of precomputing the tapers $\boldsymbol{s}_0, \ldots, \boldsymbol{s}_{K-1}$ is $O(KN \log N)$ operations, due to the fact that $\boldsymbol{B}$ commutes with a tridiagonal matrix [5].

In [7], it is shown that if $S(f)$ is twice differentiable, then bias and variance of $\widehat{S}_K^{\text{mt}}(f)$ are bounded by

$$\text{Bias}\left(\widehat{S}_K^{\text{mt}}(f)\right) \lesssim \frac{W^2}{6} S''(f),$$
$$\text{Var}\left(\widehat{S}_K^{\text{mt}}(f)\right) \lesssim \frac{1}{K} S(f)^2,$$

and thus, the mean-squared error is bounded by

$$\text{MSE}\left(\widehat{S}_K^{\text{mt}}(f)\right) \lesssim \frac{W^4}{36} S''(f)^2 + \frac{1}{K} S(f)^2.$$

Since $K \approx 2NW$, this bound is minimized when

$$W \sim \left[\frac{9S(f)}{2S''(f)}\right]^{2/5} N^{-1/5} \quad \text{and} \quad K \sim \left[\frac{12S(f)}{S''(f)}\right]^{2/5} N^{4/5}.$$

Similar analysis is done in [4] for sinusoidal tapers and in [1] for Slepian tapers. In general, fewer tapers are used for more rapidly varying spectra, but for any fixed spectrum $S(f)$ and for large N, the optimal number of tapers is $K = O(N^{4/5})$. However, this choice requires precomputing $O(N^{4/5})$ tapers and then $O(N^{4/5})$ length-N FFTs to evaluate $\widehat{S}_K^{(\cdot)}(f)$ at all N grid frequencies. This involves $O(N^{9/5} \log N)$ operations.

In this work, we present approximations $\widetilde{S}_K^{\text{mt}}(f)$ and $\widetilde{S}_K^{\text{eig}}(f)$ to $\widehat{S}_K^{\text{mt}}(f)$ and $\widehat{S}_K^{\text{eig}}(f)$ respectively which satisfy

$$\left| \widehat{S}_K^{(\cdot)}(f) - \widetilde{S}_K^{(\cdot)}(f) \right| \le \frac{O(\epsilon)}{K} \|\boldsymbol{x}\|_2^2 \quad \text{for all } f \in \mathbb{R},$$

and which can be evaluated at all grid frequencies in $O(N \log^2 N \log \frac{1}{\epsilon})$ operations. Also, the required precomputation for these approximations takes only $O(N \log^2 N \log \frac{1}{\epsilon})$ operations. When the number of tapers is $K \gtrsim \log N \log \frac{1}{\epsilon}$, evaluating $\widetilde{S}_K^{(\cdot)}(f)$ at the N grid frequencies will be significantly faster than evaluating $\widehat{S}_K^{(\cdot)}(f)$ at the N grid frequencies.

II. Intermediate Results

A. Fast algorithm for computing $\widehat{S}_N^{\text{eig}}(f)$

To begin developing our fast approximations for $\widehat{S}_K^{\text{mt}}(f)$ and $\widehat{S}_K^{\text{eig}}(f)$, we first consider the eigenvalue weighted multitaper spectral estimate with N tapers instead of $K \approx 2NW$, i.e.,[2]

$$\widehat{S}_N^{\text{eig}}(f) = \frac{1}{2NW} \sum_{k=0}^{N-1} \lambda_k \widehat{S}_k(f).$$

Using an eigendecomposition, we can write $\boldsymbol{B} = \boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^*$, where $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}_0 & \cdots & \boldsymbol{s}_{N-1} \end{bmatrix}$ and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_0, \ldots, \lambda_{N-1})$. For any $f \in \mathbb{R}$, we let $\boldsymbol{E}_f \in \mathbb{C}^{N \times N}$ be a diagonal matrix with diagonal entries $\boldsymbol{E}_f[n,n] = e^{j2\pi f n}$. Then, $\widehat{S}_N^{\text{eig}}(f)$ satisfies

$$\begin{aligned} 2NW\ \widehat{S}_N^{\text{eig}}(f) &= \sum_{k=0}^{N-1} \lambda_k \widehat{S}_k(f) \\ &= \sum_{k=0}^{N-1} \lambda_k \left| \sum_{n=0}^{N-1} \boldsymbol{s}_k[n]\boldsymbol{x}[n]e^{-j2\pi f n} \right|^2 \\ &= \sum_{k=0}^{N-1} \lambda_k \left| \boldsymbol{s}_k^* \boldsymbol{E}_f^* \boldsymbol{x} \right|^2 \\ &= \boldsymbol{x}^* \boldsymbol{E}_f \boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^* \boldsymbol{E}_f^* \boldsymbol{x} \\ &= \boldsymbol{x}^* \boldsymbol{E}_f \boldsymbol{B} \boldsymbol{E}_f^* \boldsymbol{x}. \end{aligned}$$

This gives us a a formula for $\widehat{S}_N^{\text{eig}}(f)$ which does not require computing any of the Slepian tapers. Since $\boldsymbol{B}$ is a Toeplitz matrix, it can be extended to a circulant matrix, which is diagonalized by an FFT matrix. Using this fact, we can get an alternate formula for $\widehat{S}_N^{\text{eig}}(\frac{m}{N})$ for all $m \in [N]$ as follows.

First we define a vector of sinc samples

$$\boldsymbol{b}[\ell] = \begin{cases} \dfrac{\sin[2\pi W \ell]}{\pi \ell} & \ell \in [N], \\ 0 & \ell = N, \\ \dfrac{\sin[2\pi W (2N-\ell)]}{\pi(2N-\ell)} & \ell \in [2N] \setminus [N+1], \end{cases}$$

a zeropadding matrix

$$\boldsymbol{Z} = \begin{bmatrix} \mathbf{I}_{N\times N} \\ \mathbf{0}_{N \times N} \end{bmatrix},$$

a length-$2N$ FFT matrix defined by

$$\boldsymbol{F}[m,n] = e^{-j\pi m n/N} \quad \text{for } m,n \in [2N],$$

and a vector

$$\boldsymbol{y} = \boldsymbol{F}^{-1}\left(\boldsymbol{b} \circ \boldsymbol{F} \left|\boldsymbol{F}\boldsymbol{Z}\boldsymbol{x}\right|^2\right),$$

where we use the notation $\circ$ to be the elementwise product, i.e., $(\boldsymbol{p} \circ \boldsymbol{q})[\ell] = \boldsymbol{p}[\ell]\boldsymbol{q}[\ell]$, and $|\ |^2$ to denote the elementwise magnitude-squared, i.e., $(|\boldsymbol{p}|^2)[\ell] = |\boldsymbol{p}[\ell]|^2$.

With these definitions, we have $\widehat{S}_N^{\text{eig}}(\frac{m}{N}) = \frac{1}{2NW}\boldsymbol{y}[2m]$ for all $m \in [N]$. The derivation of this fact is deferred to a future publication. Computing $\boldsymbol{y} = \boldsymbol{F}^{-1}\left(\boldsymbol{b} \circ \boldsymbol{F}|\boldsymbol{F}\boldsymbol{Z}\boldsymbol{x}|^2\right)$ can be done in $O(N \log N)$ operations via three length-$2N$ FFTs and a few pointwise multiplications of length-$2N$ vectors. Then, we can obtain $\widehat{S}_N^{\text{eig}}(\frac{m}{N}) = \frac{1}{2NW}\boldsymbol{y}[2m]$ for $m \in [N]$ by downsampling and scaling $\boldsymbol{z}$.

B. Approximations for General Multitaper Spectral Estimates

Next, we present a lemma regarding approximations to spectral estimates which use orthonormal tapers.

Lemma 1. *Let $\boldsymbol{x} \in \mathbb{C}^N$ be a vector of N equispaced samples, and let $\{\boldsymbol{v}_k\}_{k=0}^{N-1}$ be any orthonormal set of tapers in $\mathbb{C}^N$. For each $k \in [N]$, define a tapered spectral estimate*

$$V_k(f) = \left| \sum_{n=0}^{N-1} \boldsymbol{v}_k[n]\boldsymbol{x}[n]e^{-j2\pi f n} \right|^2.$$

Also, let $\{\gamma_k\}_{k=0}^{N-1}$ and $\{\widetilde{\gamma}_k\}_{k=0}^{N-1}$ be real coefficients, and then define a multitaper spectral estimate $\widehat{V}(f)$ and an approximation $\widetilde{V}(f)$ by

$$\widehat{V}(f) = \sum_{k=0}^{N-1} \gamma_k V_k(f) \quad \text{and} \quad \widetilde{V}(f) = \sum_{k=0}^{N-1} \widetilde{\gamma}_k V_k(f).$$

Then, for any frequency $f \in \mathbb{R}$, we have

$$\left| \widehat{V}(f) - \widetilde{V}(f) \right| \le \left(\max_k |\gamma_k - \widetilde{\gamma}_k| \right) \|x\|_2^2.$$

Proof. Let $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_0 & \cdots & \boldsymbol{v}_{N-1} \end{bmatrix}$, and let $\boldsymbol{\Gamma}, \widetilde{\boldsymbol{\Gamma}} \in \mathbb{R}^{N \times N}$, and $\boldsymbol{E}_f \in \mathbb{C}^{N\times N}$ be diagonal matrices whose diagonal entries are $\boldsymbol{\Gamma}[n,n] = \gamma_n$, $\widetilde{\boldsymbol{\Gamma}}[n,n] = \widetilde{\gamma}_n$, and $\boldsymbol{E}_f[n,n] = e^{j2\pi f n}$ for $n \in [N]$. Then, using a similar argument as used to show that $2NW\ \widehat{S}_N^{\text{eig}}(f) = \boldsymbol{x}^*\boldsymbol{E}_f\boldsymbol{B}\boldsymbol{E}_f^*\boldsymbol{x}$, one can show that

$$\widehat{V}(f) = \boldsymbol{x}^*\boldsymbol{E}_f\boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{V}^*\boldsymbol{E}_f^*\boldsymbol{x}$$

and

$$\widetilde{V}(f) = \boldsymbol{x}^*\boldsymbol{E}_f\boldsymbol{V}\widetilde{\boldsymbol{\Gamma}}\boldsymbol{V}^*\boldsymbol{E}_f^*\boldsymbol{x}.$$

Since $\boldsymbol{V}$ is orthonormal, $\|\boldsymbol{V}\| = \|\boldsymbol{V}^*\| = 1$. Since $\boldsymbol{E}_f$ is diagonal, and all the diagonal entries have modulus 1, $\|\boldsymbol{E}_f\| = \|\boldsymbol{E}_f^*\| = 1$. Hence, for any $f \in \mathbb{R}$, we can bound

$$\begin{aligned} \left| \widehat{V}(f) - \widetilde{V}(f) \right| &= \left| \boldsymbol{x}^*\boldsymbol{E}_f\boldsymbol{V}\left(\boldsymbol{\Gamma} - \widetilde{\boldsymbol{\Gamma}}\right)\boldsymbol{V}^*\boldsymbol{E}_f^*\boldsymbol{x} \right| \\ &\le \|\boldsymbol{x}\|_2 \|\boldsymbol{E}_f\| \|\boldsymbol{V}\| \|\boldsymbol{\Gamma} - \widetilde{\boldsymbol{\Gamma}}\| \|\boldsymbol{V}^*\| \|\boldsymbol{E}_f^*\| \|\boldsymbol{x}\|_2 \\ &= \left(\max_k |\gamma_k - \widetilde{\gamma}_k| \right) \|x\|_2^2, \end{aligned}$$

as desired. □

[2]Here, we have used the fact that $\sum_{k=0}^{N-1} \lambda_k = \operatorname{tr} \boldsymbol{B} = 2NW$.

C. Prolate matrix eigenvalue behavior

The eigenvalues $\lambda_0 > \lambda_1 > \cdots > \lambda_{N-1}$ of $\boldsymbol{B}$ are all strictly between 0 and 1, and they have a clustering behavior. For fixed $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, \frac{1}{2})$ and large N, slightly less than $2NW$ eigenvalues are between $1-\epsilon$ and 1, slightly less than $N - 2NW$ eigenvalues are between 0 and ϵ, and very few eigenvalues are between ϵ and $1-\epsilon$. In [2], it is shown that for fixed $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, \frac{1}{2})$,

$$\#\left\{k : \epsilon < \lambda_k < 1-\epsilon\right\} \sim \frac{2}{\pi^2} \log N \log\left(\frac{1}{\epsilon} - 1\right)$$

as $N \to \infty$. Also, for any $N \in \mathbb{N}$, $W \in (0, \frac{1}{2})$ and $\epsilon \in (0, \frac{1}{2})$,

$$\#\left\{k : \epsilon < \lambda_k < 1-\epsilon\right\} \le \left(\frac{8}{\pi^2}\log(8N) + 12\right) \log\left(\frac{15}{\epsilon}\right).$$

In the subsections that follow, we assume that for a given $N \in \mathbb{N}$ and $W \in (0, \frac{1}{2})$, the parameters K and $\epsilon \in (0, \frac{1}{2})$ are chosen such that $\lambda_{K-1} \ge \frac{1}{2}$ and $\lambda_K \le 1-\epsilon$. This restriction only forces K to be slightly less than $2NW$. We then partition the indices $[N]$ into four sets

$$\begin{aligned}
\mathcal{I}_1 &= \{k \in [K] : \lambda_k \ge 1-\epsilon\},\\
\mathcal{I}_2 &= \{k \in [K] : \epsilon < \lambda_k < 1-\epsilon\},\\
\mathcal{I}_3 &= \{k \in [N] \setminus [K] : \epsilon < \lambda_k < 1-\epsilon\},\\
\mathcal{I}_4 &= \{k \in [N] \setminus [K] : \lambda_k \le \epsilon\}.
\end{aligned}$$

From the above theory, we have that $\#(\mathcal{I}_2 \cup \mathcal{I}_3) = \#\{k : \epsilon < \lambda_k < 1-\epsilon\} = O(\log N \log \frac{1}{\epsilon})$. Hence, it is possible to precompute the eigenvalues λ_k and DPSS tapers $\boldsymbol{s}_k$ for all $k \in \mathcal{I}_2 \cup \mathcal{I}_3$ in $O(N \log^2 N \log \frac{1}{\epsilon})$ operations. In the following subsections, we will assume that we have precomputed λ_k and $\boldsymbol{s}_k$ for all $k \in \mathcal{I}_2 \cup \mathcal{I}_3$, but not for any $k \in \mathcal{I}_1 \cup \mathcal{I}_4$.

III. Fast Approximations

A. Fast algorithm for approximating $\widehat{S}^{\text{mt}}_K(f)$

The unweighted multitaper spectral estimate $\widehat{S}^{\text{mt}}_K(f)$ is given by

$$\widehat{S}^{\text{mt}}_K(f) = \frac{1}{K}\sum_{k=0}^{K-1} \widehat{S}_k(f) = \sum_{k \in \mathcal{I}_1 \cup \mathcal{I}_2} \frac{1}{K} \widehat{S}_k(f).$$

We then define an approximation by

$$\begin{aligned}
\widetilde{S}^{\text{mt}}_K(f) &:= \frac{2NW}{K}\widehat{S}^{\text{eig}}_N(f) + \sum_{k\in\mathcal{I}_2} \frac{1-\lambda_k}{K}\widehat{S}_k(f) - \sum_{k\in\mathcal{I}_3}\frac{\lambda_k}{K}\widehat{S}_k(f)\\
&= \sum_{k=0}^{N-1}\frac{\lambda_k}{K}\widehat{S}_k(f) + \sum_{k\in\mathcal{I}_2} \frac{1-\lambda_k}{K}\widehat{S}_k(f) - \sum_{k\in\mathcal{I}_3}\frac{\lambda_k}{K}\widehat{S}_k(f)\\
&= \sum_{k\in\mathcal{I}_1\cup\mathcal{I}_4}\frac{\lambda_k}{K}\widehat{S}_k(f) + \sum_{k\in\mathcal{I}_2}\frac{1}{K}\widehat{S}_k(f)
\end{aligned}$$

Thus, $\widehat{S}^{\text{mt}}_K(f)$ and $\widetilde{S}^{\text{mt}}_K(f)$ can be written as

$$\widehat{S}^{\text{mt}}_K(f) = \sum_{k=0}^{N-1}\gamma^{\text{mt}}_k \widehat{S}_k(f) \quad \text{and} \quad \widetilde{S}^{\text{mt}}_K(f) = \sum_{k=0}^{N-1}\widetilde{\gamma}^{\text{mt}}_k \widehat{S}_k(f)$$

where

$$\gamma^{\text{mt}}_k = \begin{cases} \frac{1}{K} & k \in \mathcal{I}_1 \cup \mathcal{I}_2,\\ 0 & k \in \mathcal{I}_3\cup\mathcal{I}_4, \end{cases} \quad \text{and} \quad \widetilde{\gamma}^{\text{mt}}_k = \begin{cases} \frac{\lambda_k}{K} & k \in \mathcal{I}_1\cup\mathcal{I}_4,\\ \frac{1}{K} & k\in\mathcal{I}_2,\\ 0 & k \in \mathcal{I}_3. \end{cases}$$

We now consider $\text{gap}^{\text{mt}}_k := |\gamma^{\text{mt}}_k - \widetilde{\gamma}^{\text{mt}}_k|$. For $k \in \mathcal{I}_1$, we have $\lambda_k \ge 1-\epsilon$, and thus,

$$\text{gap}^{\text{mt}}_k = \left|\frac{1}{K} - \frac{\lambda_k}{K}\right| = \frac{1-\lambda_k}{K} \le \frac{\epsilon}{K}.$$

For $k \in \mathcal{I}_2 \cup \mathcal{I}_3$ we have $\gamma^{\text{mt}}_k = \widetilde{\gamma}^{\text{mt}}_k$, i.e., $\text{gap}^{\text{mt}}_k = 0$. For $k \in \mathcal{I}_4$, we have $\lambda_k \le \epsilon$, and thus,

$$\text{gap}^{\text{mt}}_k = \left|0 - \frac{\lambda_k}{K}\right| = \frac{\lambda_k}{K} \le \frac{\epsilon}{K}.$$

Hence, $\text{gap}^{\text{mt}}_k \le \dfrac{\epsilon}{K}$ for all $k \in [N]$, and thus by Lemma 1, we have

$$\left|\widehat{S}^{\text{mt}}_K(f) - \widetilde{S}^{\text{mt}}_K(f)\right| \le \frac{\epsilon}{K}\|\boldsymbol{x}\|_2^2.$$

Finally, evaluating the approximation

$$\widetilde{S}^{\text{mt}}_K(f) := \frac{2NW}{K}\widehat{S}^{\text{eig}}_N(f) + \sum_{k\in\mathcal{I}_2}\frac{1-\lambda_k}{K}\widehat{S}_k(f) - \sum_{k\in\mathcal{I}_3}\frac{\lambda_k}{K}\widehat{S}_k(f)$$

at the N grid frequencies requires evaluating $\widehat{S}^{\text{eig}}_N(f)$ and $\widehat{S}_k(f)$ for all $k \in \mathcal{I}_2 \cup \mathcal{I}_3$ at the N grid frequencies. Evaluating $\widehat{S}^{\text{eig}}_N(f)$ at the grid frequencies takes $O(N\log N)$ operations, as shown in Section II-A. For each $k \in \mathcal{I}_2 \cup \mathcal{I}_3$, evaluating $\widehat{S}_k(f)$ at the grid frequencies takes $O(N\log N)$ operations. Since $\#(\mathcal{I}_2 \cup \mathcal{I}_3) = O(\log N \log\frac{1}{\epsilon})$, the total computation required is $O(N\log^2 N \log\frac{1}{\epsilon})$ operations.

B. Fast algorithm for approximating $\widehat{S}^{\text{eig}}_K(f)$

The eigenvalue weighted multitaper spectral estimate $\widehat{S}^{\text{eig}}_K(f)$ is given by:

$$\widehat{S}^{\text{eig}}_K(f) = \frac{\sum_{k=0}^{K-1}\lambda_k\widehat{S}_k(f)}{\sum_{k=0}^{K-1}\lambda_k} = \sum_{k\in\mathcal{I}_1\cup\mathcal{I}_2}\frac{\lambda_k}{\Sigma_K}\widehat{S}_k(f),$$

where

$$\Sigma_K := \sum_{k=0}^{K-1}\lambda_k = \sum_{k\in\mathcal{I}_1}\lambda_k + \sum_{k\in\mathcal{I}_2}\lambda_k.$$

We then define an approximation by

$$\begin{aligned}
\widetilde{S}^{\text{eig}}_K(f) :&= \frac{2NW}{\widetilde{\Sigma}_K}\widehat{S}^{\text{eig}}_N(f) - \frac{1}{\widetilde{\Sigma}_K}\sum_{k\in\mathcal{I}_3}\lambda_k\widehat{S}_k(f)\\
&= \frac{1}{\widetilde{\Sigma}_K}\sum_{k=0}^{N-1}\lambda_k\widehat{S}_k(f) - \frac{1}{\widetilde{\Sigma}_K}\sum_{k\in\mathcal{I}_3}\lambda_k\widehat{S}_k(f)\\
&= \sum_{k\notin\mathcal{I}_3}\frac{\lambda_k}{\widetilde{\Sigma}_K}\widehat{S}_k(f)
\end{aligned}$$

where

$$\widetilde{\Sigma}_K := K - \sum_{k\in\mathcal{I}_2}(1-\lambda_k) = \sum_{k\in\mathcal{I}_1} 1 + \sum_{k\in\mathcal{I}_2}\lambda_k.$$

Thus, $\widehat{S}_K^{\text{eig}}(f)$ and $\widetilde{S}_K^{\text{eig}}(f)$ can be written as

$$\widehat{S}_K^{\text{eig}}(f) = \sum_{k=0}^{N-1} \gamma_k^{\text{eig}} \widehat{S}_k(f) \quad \text{and} \quad \widetilde{S}_K^{\text{eig}}(f) = \sum_{k=0}^{N-1} \widetilde{\gamma}_k^{\text{eig}} \widehat{S}_k(f)$$

where

$$\gamma_k^{\text{eig}} = \begin{cases} \frac{\lambda_k}{\Sigma_K} & k \in \mathcal{I}_1 \cup \mathcal{I}_2, \\ 0 & k \in \mathcal{I}_3 \cup \mathcal{I}_4, \end{cases} \quad \text{and} \quad \widetilde{\gamma}_k^{\text{eig}} = \begin{cases} \frac{\lambda_k}{\widetilde{\Sigma}_K} & k \notin \mathcal{I}_3, \\ 0 & k \in \mathcal{I}_3. \end{cases}.$$

To bound $\text{gap}_k^{\text{eig}} := \left|\gamma_k^{\text{eig}} - \widetilde{\gamma}_k^{\text{eig}}\right|$, we first note that

$$0 \le \widetilde{\Sigma}_K - \Sigma_K = \sum_{k \in \mathcal{I}_1} (1 - \lambda_k) \le \epsilon \#(\mathcal{I}_1) \le K\epsilon,$$

and

$$\widetilde{\Sigma}_K \ge \Sigma_K = \sum_{k=0}^{K-1} \lambda_k \ge \sum_{k=0}^{K-1} \lambda_{K-1} = K\lambda_{K-1} \ge \frac{K}{2}.$$

For $k \in \mathcal{I}_1 \cup \mathcal{I}_2$, we have $0 \le \lambda_k \le 1$, and thus,

$$\text{gap}_k^{\text{eig}} = \left|\frac{\lambda_k}{\Sigma_K} - \frac{\lambda_k}{\widetilde{\Sigma}_K}\right| = \frac{\lambda_k(\widetilde{\Sigma}_K - \Sigma_K)}{\widetilde{\Sigma}_K \Sigma_K} \le \frac{1 \cdot K\epsilon}{(\frac{K}{2})^2} \le \frac{4\epsilon}{K}.$$

For $k \in \mathcal{I}_3$ we have $\gamma_k^{\text{eig}} = \widetilde{\gamma}_k^{\text{eig}} = 0$, i.e., $\text{gap}_k^{\text{eig}} = 0$. For $k \in \mathcal{I}_4$, we have $\lambda_k \le \epsilon$, and thus,

$$\text{gap}_k^{\text{eig}} = \left|0 - \frac{\lambda_k}{\widetilde{\Sigma}_K}\right| = \frac{\lambda_k}{\widetilde{\Sigma}_K} \le \frac{\epsilon}{\frac{K}{2}} = \frac{2\epsilon}{K} \le \frac{4\epsilon}{K}.$$

Hence, $\text{gap}_k^{\text{eig}} \le \frac{4\epsilon}{K}$ for all $k \in [N]$, and thus by Lemma 1, we have

$$\left|\widehat{S}_K^{\text{eig}}(f) - \widetilde{S}_K^{\text{eig}}(f)\right| \le \frac{4\epsilon}{K} \|\boldsymbol{x}\|_2^2.$$

Finally, evaluating the approximation

$$\widetilde{S}_K^{\text{eig}}(f) := \frac{2NW}{\widetilde{\Sigma}_K} \widehat{S}_N^{\text{eig}}(f) - \frac{1}{\widetilde{\Sigma}_K} \sum_{k \in \mathcal{I}_3} \lambda_k \widehat{S}_k(f)$$

at the N grid frequencies can be done in $O(N \log^2 N \log \frac{1}{\epsilon})$ operations in a similar manner as can be done for $\widetilde{S}_K^{\text{mt}}(f)$.

IV. Simulations

To test our fast method for multitaper spectral estimation, we first generate $N = 2^{20}$ samples of an ARMA$(12, 8)$ process. We then try the following methods of spectral estimation:

1) Thomson's unweighted multitaper method with $W = 3.6 \times 10^{-5}$ ($2NW \approx 75.5$), and $K = 63$ tapers.
2) Our fast approximation to Thomson's unweighted multitaper method with $W = 2.7 \times 10^{-3}$ ($2NW \approx 5662.3$), $K = 5641$ tapers, and an approximation parameter of $\epsilon = 10^{-12}$.

Note that for both methods, the number of tapers K was chosen such that $\lambda_{K-1} > 1 - 10^{-9} > \lambda_K$, which severely reduces the broadband bias of the tapered estimates. This is necessary due to the high dynamic range of the true spectrum. For the first method, the half-bandwidth parameter $W = 2.7 \times 10^{-3}$ was chosen according to the optimal number of tapers suggested in [7]. For the second method, the half-bandwidth parameter $W = 3.6 \times 10^{-5}$ was chosen so that both methods run in a comparable amount of time.

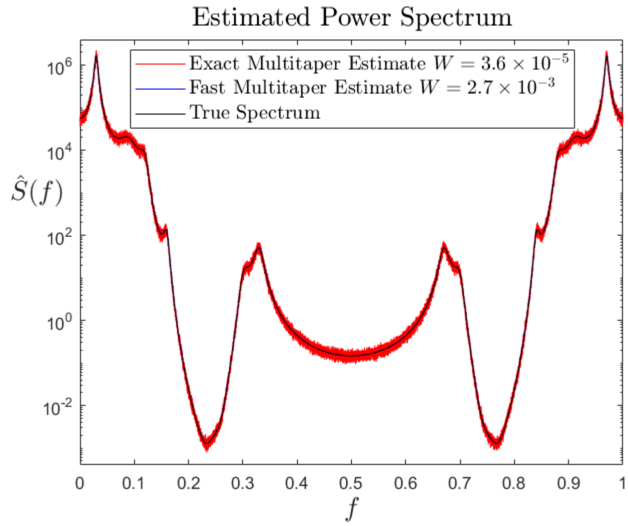

Fig. 1. Plots of the spectrum of the ARMA$(12, 8)$ process, and the two estimates of this spectrum.

A plot the exact power spectrum of the ARMA$(12, 8)$ process and the estimated spectra are shown in Figure 1. The precomputation time, run time, and root-mean-squared-logarithmic errors are shown in the table below. Both methods run in approximately the same amount of time due to the fact that our fast approximation only needed to compute $\#(\mathcal{I}_2 \cup \mathcal{I}_3) = 56$ Slepian tapers. However, the fast approximation has greater accuracy due to the fact that it approximates a multitaper estimate with $K = 5641$ tapers.

Method	Precomputation time	Time	RMSLE
1	28.15 s	2.989 s	0.5498 dB
2	25.33 s	2.932 s	0.1602 dB

Acknowledgements

This work was supported, in part, by NSF grants CCF-1350616 and CCF-1409406, a grant from Lockheed Martin, and a gift from the Alfred P. Sloan Foundation.

References

[1] L. Abreu and J. Romero, "MSE Estimates for Multitaper Spectral Estimation and Off-Grid Compressive Sensing," *IEEE Trans. Inform. Theory*, vol. 63, no. 12, pp. 7770–7776, 2017.

[2] S. Karnik, Z. Zhu, M. Wakin, J. Romberg, & M. Davenport "The fast Slepian transform," to appear in *Appl. Comput. Harm. Anal.*, 2019.

[3] G. Prieto, R. Parker, D. Thomson, F. Vernon, and R. Graham, "Reducing the bias of multitaper spectrum estimates," *Geophys. J. Int.*, vol. 171, no. 3, pp. 1269–1281, 2007.

[4] K. Riedel and A. Sidorenko, "Minimum bias multiple taper spectral estimation," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 188–195, 1995.

[5] 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, 1978.

[6] D. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1095, 1982.

[7] A. Walden, "A unified view of multitaper multivariate spectral estimation," *Biometrika*, vol. 87, no. 4, pp. 767–788, 2000.