Smoothing span selection for the multitaper spectral estimate in the analysis of EEG data

Deborah Schneider-Luftman, *not in IEEE yet*

I. INTRODUCTION

should show up in desktop branch

In the recording Electroencephalograms (EEG), neurological signals are simultaneously obtained from different regions of the brain and give rise to data in the form of a multivariate time series. In this context, it is particularly relevant to analyze this data in the frequency domain, rather than the time domain. Meaningful results can be extracted from the Spectral density function (SDF) matrix of the data, which can be viewed as the complex-valued covariance matrix of the Fourier-transformed data.

The aim is to retrieve the spectral matrix, at all frequencies, for each patient in order to highlight differences between positive and negative syndrome patients, using the control individuals as benchmarks. The multitaper estimate is a strong candidate for the estimation of this quantity. Not only does it have good bias and variance properties in finite samples, it is also very effective in controlling spectral leakage [1].

Much like in the smoothed peridogram, the number of tapers K used in its constructions acts as a smoothing span. Thus it must be carefully selected, and reflect a balance between bias/variance reduction and good spectral resolution.

While there exists many smoothing span selection schemes for the smoothed periodogram, little research has been done in the case of the multitaper estimate, in spite of its potential for application to analysis of neurological data.

In this paper we propose data-driven schemes that select a value for K that optimizes specific target criteria discussed in section II. The schemes are then applied, first to data simulated from a VAR(1) model, then to the data sampled from the patients of all three groups in section III on page 3.

II. METHODS

EEG data set

The data set is the same as the one discussed in [2]: EEG are taken from 34 mental health patients (all male, aged between 20 and 60 years) from the Serbsky Institute in Moscow. Of these patients, 15 were diagnosed with positive syndrome schizophrenia, 19 were diagnosed with negative syndrome schizophrenia. In addition to this, a control group of 24 healthy subjects of similar age to the patients is introduced to the study. We denote the positive syndrome group as "Positive", the negative syndrome group as "Negative" and the control group as "Control".

Department of Mathematics, Imperial College London.

The EEGs are recorded at a sampling rate of 100Hz ($\Delta t = 0.01$s, Nyquist frequency $f_N = \frac{1}{2\Delta t} = 50$Hz) from $p = 10$ scalp sites, named F3, F4, C3, C4, T3, T4, P3, P4, O1 and O2. The raw signals are then put thought a 6Hz Butterworth filter with bandpass 0.5-45 Hz, and subsequently down-sampled by a factor of 5. The resulting signal is a multivariate time series $\{X_{j,t}\}$ with dimension $p = 10$ where $j = 1, ..., 10$ represents the re-labeling each scalp site:

1	2	3	4	5	6	7	8	9	10
F3	F4	C3	C4	T3	T4	P3	P4	O1	O2

Spectral matrix estimation

In this context we work with the spectral matrix $\mathbf{S}(f)$ of the multivariate time series $\{X_{i,t}\}$, $i \leq p$, $t \in [0, N-1]$, where each p-dimensional vector $\mathbf{X}_t$ is viewed as a zero-mean and strongly stationary processes $X_{m,t}$, $\forall m \leq p$. The true spectral matrix $\mathbf{S}(f) = \{S(f)_{lm}\}_{l,m=1}^p$ is given by:

$$S(f)_{lm} = \Delta t \sum_{\tau=-\infty}^{+\infty} s(\tau)_{lm} e^{-i2\pi f \tau \Delta t}$$

for each frequency $f \in [-f_N, f_N]$, f_N being the Nyquist frequency $f_N = \frac{1}{2\Delta t}$ and Δt the sample time interval. The term:

$$\forall t \leq T, \ s(\tau)_{lm} = \mathbb{E}(X_{l,t+\tau} X_{m,t})$$

is the cross-covariance between the processes $\mathbf{X}_l$ and $\mathbf{X}_m \in \{\mathbf{X}_i\}_i^{N-1}$, $\mathbf{X}_i \in \mathbb{R}^p$. $\forall i \leq N-1$.

The matrix $\mathbf{S}(f)$ is traditionally estimated via the Periodogram:

$$\mathbf{J}(f) = \frac{\Delta t^{-1/2}}{\sqrt{N}} \sum_{t=0}^{N-1} \mathbf{X}_t e^{-i2\pi f t \Delta t} \in \mathbb{C}^p,$$

$$\hat{\mathbf{S}}(f) = \mathbf{J}(f)\mathbf{J}(f)^H \in \mathbb{C}^{p \times p}.$$

This estimator however is known to be biased asymptotically and to have stability and inversion issues. Instead, we use an alternative estimation procedure, similar to the smoothed periodogram, called the multitaper spectral estimate [1].

We start with tapers $\{h_{k,t} | t = 0, ..., N-1\}$ for $k = 0, ..., K-1$, where $h_{k,t}$, constants in $\mathbb{R}$ and orthogonal, displayed below ([2]):

$$\sum_t h_{j,t} h_{k,t} = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{otherwise} \end{cases} \tag{II.1}$$

We introduce a Fourier transform on $\mathbf{X}$:

$$\mathbf{J}_k(f) = \Delta t^{-1/2} \sum_t h_{k,t} \mathbf{X}_t e^{-i2\pi f t \Delta t} \in \mathbb{C}^p. \tag{II.2}$$

From this we can create a multitaper estimate for $\mathbf{S}(f)$:

$$\hat{\mathbf{S}}^{(mt)}(f) = \frac{1}{K}\sum_{k=0}^{K-1}\mathbf{J}_k(f)\mathbf{J}_k^H(f) \in \mathbb{C}^{p\times p} \qquad \text{(II.3)}$$

There are several types of tapers $\{h_{k,t}\}$ that satisfy the relation in equation (II.1) on the preceding page ([1]), the most widely used ones being the Slepian tapers [3] and the Sine tapers [4]. Using these, we can create the estimator in equation (II.3) and obtain strong bias, variance properties, as well as reduced broadband bias [5].

Number of tapers K as a design parameter

To maximise the beneficial properties of the estimator in equation (II.3), the number of tapers K used in its construction must be carefully selected. The bias and variance of $\hat{\mathbf{S}}^{(mt)}(f)$ are inversely related to K for all frequencies $f \in (0, f_N)$

$$\mathbb{V}(\hat{\mathbf{S}}^{(mt)}(f)) = \mathbf{S}(f)^2/K$$
$$\mathbb{E}(\hat{\mathbf{S}}^{(mt)}(f)) \simeq \mathbf{S}(f) + \mathcal{O}(1/K)$$

Hence, it is best to set K to be relatively large. However, K cannot be set arbitrarily large, as it is directly related to the bandwidth of the resulting estimator $\hat{\mathbf{S}}^{(mt)}(f))$. Denote W the desired half-bandwidth, and B the effective bandwidth [6]:

$$\text{Slepian tapers: } K << 2W - 1,\ B = 2W$$
$$\text{Sine tapers: } B = \frac{K+1}{(N+1)}$$

The bandwidth B of the estimate $\hat{\mathbf{S}}^{(mt)}(f)$ directly impacts its quality: with a bandwidth B set too high, much of the information in the initial time series data is lost in the frequency domain. In all the existing literature on multitaper spectral analysis, the effective bandwidth B is controlled by the desired bandwidth W, which is input at the start as a design parameter. So far, there is no specific indication on how W - and therefore K - should be set. We thus need to formulate a selection procedure that aims to select a number of taper K such that the resulting estimator $\hat{\mathbf{S}}^{(mt)}$ has as little bias and variance as possible, while not setting B too high.

Selection criteria

A natural approach to is to extend the results previously developed for the selection of the smoothing span in the smoothed periodogram. The method exposed in [7] aims to minimise the Mean Integrated squared Error (MISE) of $\hat{\mathbf{S}}_K^{(mt)}$ in HS norm

$$
\begin{aligned}
K^* &= \operatorname*{argmin}_K \mathcal{MISE}_{HS}(\hat{\mathbf{S}}_K^{(mt)}, \mathbf{S}) \\
&= \operatorname*{argmin}_K \int_0^{f_N} \mathcal{R}_{HS}(\hat{\mathbf{S}}_K^{(mt)}(f), \mathbf{S}(f))df \\
&= \operatorname*{argmin}_K \mathbb{E}\left(\int_0^{f_N} \operatorname{tr}[\hat{\mathbf{S}}_K^{(mt)}(f) - \mathbf{S}(f)]^2 df\right) \\
&\simeq \operatorname*{argmin}_K \sum_{j=0}^{N_{freq}-1} \mathbb{E}\left(\operatorname{tr}[\hat{\mathbf{S}}_K^{(mt)}(f_j) - \mathbf{S}(f_j)]^2\right)
\end{aligned}
$$

The last line represents a Trapezoidal scheme approximating $\mathcal{MISE}_{HS}(\hat{\mathbf{S}}_K^{(mt)}, \mathbf{S})$ over a finite set of frequencies, where N_{freq} is the number of discretized frequencies within $(0, f_N)$, typically,

$$N_{freq} = N + 1$$

But in the context of data analysis, the quantity $\mathbf{S}(.)$ is unknown. Thus, we need to redevelop this methodology for it work solely with available data. This can be easily achieve by exploiting the independence structure of the sub-vectors $\mathbf{J}_k(f)$, with a leave-one-out cross validation [8], [9]:

$$\hat{\mathbf{S}}_{\backslash k}^{(mt)}(f) = \frac{1}{K}\sum_{\substack{u=0 \\ u\neq k}}^{K-1}\mathbf{J}_u(f)\mathbf{J}_u^H(f),$$

$$\hat{\mathcal{R}}_{HS}(\hat{\mathbf{S}}_K^{(mt)}(f), \mathbf{S}(f)) = \sum_{k=0}^{K-1}\operatorname{tr}[\hat{\mathbf{S}}_{\backslash k}^{(mt)}(f_j) - \mathbf{J}_k(f_j)\mathbf{J}_k^H(f_j)]^2,$$

$$\widehat{\mathcal{MISE}}_{HS}(\hat{\mathbf{S}}_K^{(mt)}, \mathbf{S}) = \sum_{j=0}^{N_{freq}-1}\hat{\mathcal{R}}_{HS}(\hat{\mathbf{S}}_K^{(mt)}(f_j), \mathbf{S}(f_j)),$$

$$\hat{K}^* = \operatorname*{argmin}_K \widehat{\mathcal{MISE}}_{HS}(\hat{\mathbf{S}}_K^{(mt)}, \mathbf{S}) \qquad \text{(II.4)}$$

Note how perfectly the multitaper estimate lends itself to this methodology.

It is a fact that [10],[8], for all $f \in (0, f_N)$,

$$\left|\hat{\mathcal{R}}_{HS}(\hat{\mathbf{S}}_K^{(mt)}(f), \mathbf{S}(f)) - \mathcal{R}_{HS}(\hat{\mathbf{S}}_K^{(mt)}(f), \mathbf{S}(f))\right| \to 0,\ N \to +\infty.$$

Then, given the asymptotic properties of the trapezoidal scheme, it follows that

$$\left|\widehat{\mathcal{MISE}}_{HS}(\hat{\mathbf{S}}_K^{(mt)}, \mathbf{S}) - \mathcal{MISE}_{HS}(\hat{\mathbf{S}}_K^{(mt)}, \mathbf{S})\right| \to 0,\ N \to +\infty.$$

Hence, $\hat{K}^*$ is a strong estimator for K^*, but it may not necessarily be the best smoothing span for the multitatper estimator.

An alternative is to use the multivariate Whittle likelihood [11] (see equation (II.6)), and create a minimization scheme for the multitaper estimate that is similar to the one exposed in [12]:

$$
\begin{aligned}
K_L &= \operatorname*{argmin}_K L(\mathbf{S}|\hat{\mathbf{S}}_K^{mt}) \\
\hat{K}_L &= \operatorname*{argmin}_K \hat{L}(\mathbf{S}|\hat{\mathbf{S}}_K^{mt}) \qquad \text{(II.5)}
\end{aligned}
$$

$$L(\mathbf{S}|\hat{\mathbf{S}}_K^{mt}) = \frac{1}{2f_N}\int_0^{f_N}\log(|\mathbf{S}(f)|) + \operatorname{tr}[\mathbf{S}^{-1}(f)\hat{\mathbf{S}}_K^{mt}(f)]df \qquad \text{(II.6)}$$

$$\hat{L}(\mathbf{S}|\hat{\mathbf{S}}_K^{mt}) = \frac{1}{2f_N}\sum_{j=0}^{N_{freq}-1}\log(|\mathbf{S}(f_j)|) + \operatorname{tr}[\mathbf{S}^{-1}(f_j)\hat{\mathbf{S}}_K^{mt}(f_j)] \qquad \text{(II.7)}$$

with $\hat{L}(.|.)$ defined in equation (II.7) on the preceding page and $\hat{S}$ represents a plug-in estimator. Following the guideline set in [7], we can set $\hat{S} = \hat{S}_{K_h}^{(mt)}$, where K_h is a pilot parameter, conveniently set as $K_h = \hat{K}^*$ from equation (II.4) on the previous page.

Remark 1. *If $p > K_h$ $\hat{S}_{K_h}^{(mt)}$ is singular. In that case, equation (II.7) on the preceding page cannot be computed. Instead, we could use a regularised version of $\hat{S}_{K_h}^{(mt)}$, shrinkage being one of the optimal options for its computational ease [8].*

Another possibility consists of estimating the bandwidth B_T of the time series data and adapt the number of tapers K around it. For each variable $i \leq p$, $B_{T,i}$ ca be estimated with a scheme exposed in [13]:

$$\hat{B}_{T,i} = \frac{\hat{s}_{i,0}^2}{2\Delta_t \sum_{\tau=-(N-1)}^{N-1}(1 - \frac{|\tau|}{N})\hat{s}_{i,\tau}^2},$$
$$\tilde{B}_{T,i} = \frac{5}{3}\hat{B}_T - \frac{1}{N\Delta_t}, \qquad \text{(II.8)}$$

where $\hat{s}_{i,\tau}^2$ - for $\tau \in [-(N-1), N-1]$ - represent the auto and cross covariance terms

$$\hat{s}_{i,\tau}^2 = \frac{1}{N}\sum_j^N X_{i,t}X_{i,t+|\tau|}^T$$

Working with uni-dimensional data ($p = 1$), the approach of [14] uses $\tilde{B}_T$ to evaluate K when using Slepian tapers:

1) Compute $\tilde{B}_T$ from the data $\{\mathbf{X}_t\}$ using equation (II.8),
2) Set the design parameter W as $W = \frac{\tilde{B}_T}{\alpha}$, $\alpha \in 2^{\mathbb{R}}$,
3) $K = 2NW - 1$.

How can this protocol can be extended to work for $p \geq 2$?

III. RESULTS

REFERENCES

[1] D. Percival and A. T. Walden, *Spectral analysis for Physical applications*. Cambridge University Press, 1993.

[2] T. Medkour, a. T. Walden, a. P. Burgess, and V. B. Strelets, "Brain connectivity in positive and negative syndrome schizophrenia." *Neuroscience*, vol. 169, no. 4, pp. 1779–88, Sep. 2010.

[3] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertaintyV: The discrete case," *Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, 1978.

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

[5] B. Babadi and E. N. Brown, "A review of multitaper spectral analysis." *IEEE transactions on bio-medical engineering*, vol. 61, no. 5, pp. 1555–64, May 2014.

[6] A. T. Walden, E. J. McCoy, and D. B. Percival, "The Effective Bandwidth of a Multitaper Spectral Estimator," *Biometrika*, vol. 82, no. 1, p. 201, Mar. 1995.

[7] T. C. Lee, "A stabilized bandwidth selection method for kernel smoothing of the periodogram," *Signal Processing*, vol. 81, no. 2, pp. 419–430, Feb. 2001.

[8] A. Walden and D. Schneider-Luftman, "Random Matrix Shrinkage Methods for Spectral Matrices and their inverses," *(pre-print)*.

[9] M. Fiecas and H. Ombao, "The generalized shrinkage estimator for the analysis of functional connectivity of brain signals," *The Annals of Applied Statistics*, vol. 5, no. 2A, pp. 1102–1125, Jun. 2011.

[10] Y. Chen, A. Wiesel, Y. Eldar, and A. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5016–5029, 2010.

[11] R. Dahlhaus, "Graphical interaction models for multivariate time series 1," *Metrika*, vol. 51, no. 2, pp. 157–172, Aug. 2000.

[12] H. Ombao, J. Raz, R. Strawderman, and R. V. Sachs, "A simple generalised crossvalidation method of span selection for periodogram smoothing," *Biometrika*, vol. 2, no. February 2000, pp. 1186–1192, 2001.

[13] a. T. Walden and R. E. White, "Estimating the statistical bandwidth of a time series," *Biometrika*, vol. 77, pp. 699–707, 1990.

[14] I. M. Rosado-Mendez, T. J. Hall, and J. a. Zagzebski, "Performance of an adaptive multitaper method for reducing coherent noise in spectral analysis of ultrasound backscattered echoes," *IEEE International Ultrasonics Symposium, IUS*, pp. 429–432, 2013.