

Multi-taper Spectral Analysis in Gravitational Wave Data Analysis

Bruce Allen¹ and Adrian Ottewill²

Received 2 July, 1999

Spectral estimation plays a significant role in gravitational wave data analysis. We provide a brief introduction to multi-taper methods which use multiple orthogonal tapers (or windows) to provide spectral estimators with excellent bias and variance properties. Multi-taper methods are also extremely powerful for the estimation and removal of sharp spectral peaks in the presence of noise such as arise due to power line harmonics or suspension resonances. We present examples of these methods using the GRASP (Gravitational Radiation Analysis and Simulation Package) software package.

KEY WORDS : Gravitational wave data analysis

1. INTRODUCTION

One of the most important areas of signal processing is that of the spectral analysis of signals. Since the early 1980s there has been a revolution in spectral analysis, due largely to a seminal paper by Thomson [7]. The current paper serves to give a very brief introduction to the multi-taper methods introduced by Thomson. Limitations of space mean that we shall not attempt to be rigorous in our presentation; for a fuller and more careful exposition of the subject the reader is recommended the excellent book by Percival and Walden [4], to which we will frequently refer.

¹ Department of Physics, University of Wisconsin — Milwaukee, PO Box 413, Milwaukee, Wisconsin 53201, USA. E-mail: ballen@dirac.phys.uwm.edu

² Department of Mathematical Physics, University College Dublin, Belfield, Dublin 4, Ireland. E-mail: ottewill@relativity.ucd.ie

We have implemented multi-taper methods as part of the GRASP (Gravitational Radiation Analysis and Simulation Package) software package [1]. These programs may be used to address issues relating to gravitational wave detection, both for detector characterisation, for example:

- The characterisation of power line harmonics, and specific suspension or pendulum modes. This can be used both for diagnostic purposes (e.g., track contamination of the data set by the 5th line harmonic at 300Hz) or to “clean up” the data (e.g., remove a known pendulum resonance).

and for detection purposes, for example:

- The method may be trivially extended to provide a high resolution multi-taper spectral estimate of the (complex-valued) cross-correlation spectrum $\tilde{x}^*(f)\tilde{y}(f)$ of the two input time-series $x(t)$ and $y(t)$, see the GRASP routine `multitaper_cross_spectrum()`, which is particularly important in the detection of a stochastic gravitational wave background.

2. SPECTRAL ESTIMATION

The power spectral density provides one of the most useful tools for understanding a time series. However, the problem of estimating the power spectral density from a finite observation is fraught with difficulty. Many techniques have been developed; below we shall consider only a class of non-parametric methods of power spectrum estimation since these are most appropriate to gravitational wave data analysis.

2.1. The spectral density function

The most intuitive approach to estimating the power spectrum of a signal is to compute the squared magnitude of the Fourier transform of a segment of the signal. Let $\{x_r\}$ be any realisation of the (discrete) stationary process $\{X_r\}$ with zero mean (if the mean is initially non-zero we work with the sequence with the mean subtracted) and for any $N > 0$ let

$$\tilde{x}^{(N)}(f) = \Delta t \sum_{r=1}^N x_r e^{-2\pi i f r \Delta t}, \quad (1)$$

where Δt is the sampling interval. Then

$$\int_{-f_{\text{ny}}}^{f_{\text{ny}}} \tilde{x}^{(N)}(f) e^{2\pi i f r \Delta t} df = \begin{cases} x_r & 0 \leq r \leq N \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $f_{\text{ny}} \equiv 1/(2\Delta t)$ is the Nyquist frequency, and so by Parseval's theorem

$$\sum_{r=1}^N |x_r|^2 = \frac{1}{\Delta t} \int_{-f_{\text{ny}}}^{f_{\text{ny}}} |\tilde{x}^{(N)}(f)|^2 df. \quad (3)$$

This equation leads to the interpretation of $|\tilde{x}^{(N)}(f)|^2/(\Delta t)$ as the contribution to the energy per unit frequency of $x_1, x_2 \dots x_N$ from frequencies near f .

If we take the limit as $N \rightarrow \infty$, the energy per unit frequency $|\tilde{x}^{(N)}(f)|^2/(\Delta t)$ diverges since the process is stationary (so its energy is infinite) but the power

$$\lim_{N \rightarrow \infty} \frac{|\tilde{x}^{(N)}(f)|^2}{N\Delta t} \quad (4)$$

is well-defined. This equation defines the power spectral density for the particular realisation of the process. To define it for the process itself we define

$$S(f) \equiv \lim_{N \rightarrow \infty} \frac{1}{N\Delta t} E(|\tilde{X}^{(N)}(f)|^2). \quad (5)$$

$S(f)$ is the (two-sided) power spectral density of the process defined on $[-f_{\text{ny}}, f_{\text{ny}}]$; for alternative (and more rigorous) definitions see Percival and Walden.

Note that as our time-series is real, $S(f) = S(-f)$. In the gravitational wave community it is often standard to define the one-sided power spectral density as

$$\begin{aligned} S_{\text{one-sided}}(f) &= 2S(f) \\ &= 2 \lim_{N \rightarrow \infty} \frac{1}{N\Delta t} E(|\tilde{X}^{(N)}(f)|^2) \quad \text{on } f \in [0, f_{\text{ny}}]. \end{aligned} \quad (6)$$

It will be more convenient for us to work with the two-sided power spectral density and the conversion is clearly trivial.

2.2. The periodogram

If we have at our disposal only a finite number N of observations $X_1, X_2, \dots X_N$ of the process it is natural to estimate the power spectral density by

$$\hat{S}_p^{(N)}(f) \equiv \frac{1}{N\Delta t} |\tilde{X}^{(N)}(f)|^2 = \frac{\Delta t}{N} \left| \sum_{r=1}^N X_r e^{-2\pi i f r \Delta t} \right|^2. \quad (7)$$

$\hat{S}_p^{(N)}(f)$ is called the *periodogram* and is defined on the interval $[-f_{\text{ny}}, f_{\text{ny}}]$.

Clearly as $N \rightarrow \infty$

$$\lim_{N \rightarrow \infty} E(\hat{S}_p^{(N)}(f)) = S(f) \quad \forall f, \tag{8}$$

that is, $\hat{S}_p^{(N)}(f)$ is an asymptotically unbiased estimator. Unfortunately, eq. (8) gives no measure of how large N must be for $E(\hat{S}_p^{(N)}(f))$ to be close to $S(f)$. It may be shown that

$$E(\hat{S}_p^{(N)}(f)) = \int_{-f_{ny}}^{f_{ny}} \mathcal{F}^{(N)}(f - f')S(f') \, df', \tag{9}$$

where $\mathcal{F}_N(f)$ is Fejér kernel

$$\mathcal{F}^{(N)}(f) = \frac{\Delta t}{N} \left| \sum_{r=1}^N e^{2\pi i f r \Delta t} \right|^2 = \frac{\Delta t}{N} \frac{\sin^2(N\pi f \Delta t)}{\sin^2(\pi f \Delta t)}. \tag{10}$$

$10 \log_{10}(\mathcal{F}^{(N)}(f))$ is plotted for $N = 32$ in Figure 1.

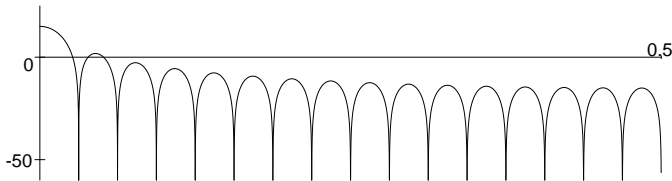


Figure 1. $10 \log_{10}(\mathcal{F}^{(N)}(f))$ for $N = 32$ and $0 \leq f \Delta t \leq 0.5$. (From its definition it is immediate that $\mathcal{F}^{(N)}(f)$ is an even function of f .)

As $N \rightarrow \infty$ the Fejér kernel converges in a distributional sense to the Dirac delta function giving rise to eq. (8). However, for finite N it has the undesirable features that it has a broad central lobe and side lobes which only die off very slowly; for illustrative purposes, $10 \log_{10}(\mathcal{F}^{(N)}(f))$ is plotted for $N = 32$ in Fig. 1. As a result eq. (9) shows that there is a great deal of contamination in $\hat{S}_p^{(N)}(f)$ due to quite distant frequencies,

i.e., there is a *leakage* of power from one part of $S(f)$ to a distant part of $\hat{S}_p^{(N)}(f)$. This is a particularly acute problem for processes with broad dynamic range where leakage from dominant frequencies can swamp the true spectral density at less dominant frequencies. This effect is not limited to small values of N , Thomson [7] cites examples where using sample sizes exceeding one million points the periodogram gives extremely biased estimates.

2.3. Tapering (windowing)

A standard solution to the problem of bias in the periodogram is that data tapering (or windowing) which was introduced by Blackman and Tukey [2]. The central idea of this approach is to replace the Fejér kernel with a kernel with reduced side-lobes.

Assume, as before, that we have at our disposal only a finite number N of observations X_1, X_2, \dots, X_N of the process. Let us weight these by some weighting factors w_1, w_2, \dots, w_N , which we will normalise to satisfy

$$\sum_{r=1}^N w_r^2 = \frac{1}{\Delta t} \int_{-f_{ny}}^{f_{ny}} |\tilde{w}(f)|^2 df = 1, \quad (11)$$

that is, that the total energy of the taper be 1. The idea of tapering is not as radical as it may seem since in our periodogram approach we are effectively weighting the observations X_1, X_2, \dots, X_N equally (with weight factor $1/\sqrt{N}$) and all other (potential) observations by weight factor 0. Indeed the sharp edges of this taper contain a lot of power at high frequency leading to significant power leakage.

The modified periodogram corresponding to the weighted series is given by

$$\hat{S}_w^{(N)}(f) \equiv \Delta t \left| \sum_{r=1}^N w_r X_r e^{-2\pi i f r \Delta t} \right|^2. \quad (12)$$

and so

$$E(\hat{S}_w^{(N)}(f)) = \int_{-f_{ny}}^{f_{ny}} \mathcal{W}^{(N)}(f - f') S(f) df, \quad (13)$$

where $\mathcal{W}^{(N)}(f)$ is given by

$$\mathcal{W}^{(N)}(f) = \Delta t \left| \sum_{r=1}^N w_r e^{2\pi i f r \Delta t} \right|^2. \quad (14)$$

The choice of normalisation (11) ensures that

$$\int_{-f_{ny}}^{f_{ny}} E(\hat{S}_w^{(N)}(f)) df = \int_{-f_{ny}}^{f_{ny}} S(f) df. \quad (15)$$

There are many standard choices for weighting factors which give a significant improvement in terms of making the central lobe as narrow as possible and making the side lobes fall off as rapidly as possible, see for example, Press et al. [5]. In general, there is a trade-off between frequency resolution and spectral variance.

2.4. Slepian tapers

Of particular importance to us is a class of tapers introduced by Slepian [6]. These are constructed by first choosing a width W for the central lobe and then demanding that the taper extremise the energy in $(-W, W)$

$$E_W \equiv \frac{1}{\Delta t} \int_{-W}^W |\tilde{w}^{(N)}(f)|^2 df, \quad (16)$$

subject to the constraint (11) that the total energy be 1. In the time domain this corresponds to extremising

$$\sum_{r=1}^N \sum_{s=1}^N w_r A_{rs} w_s \quad \text{subject to} \quad \sum_{r=1}^N w_r^2 = 1, \quad (17)$$

where

$$A_{rs} = \frac{\sin 2\pi W(r-s)}{\pi(r-s)}. \quad (18)$$

This is a standard extremisation problem whose solution is well known to be given by the eigenvectors of the real symmetric matrix A_{rs} :

$$A_{rs} w_s^{(k)} = \lambda^{(k)} w_r^{(k)}. \quad (19)$$

As A_{rs} is a real symmetric matrix, the eigenvectors corresponding to it are orthogonal so

$$\sum_{r=1}^N w_r^{(j)} w_r^{(k)} = \delta_{jk}. \quad (20)$$

In this sense we talk of the Slepian tapers as being orthogonal.

The eigenvalue $\lambda^{(k)}$ gives the confined energy E_W of the k th eigenvector $w_r^{(k)}$ and correspondingly satisfies $0 < \lambda^{(k)} \leq 1$. In fact since our

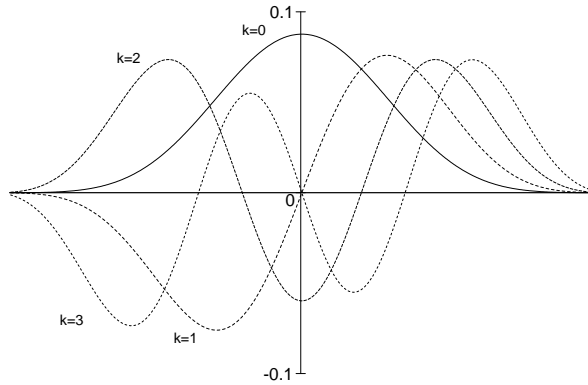


Figure 2. The first four Slepian tapers for $N = 512$, $W = 4/(N\Delta t)$. These tapers were computed by the GRASP routine `slepian_tapers()`. Note that, unlike most standard choices for windowing functions, the Slepian tapers for $k \geq 1$ assign *negative* weights to some points.

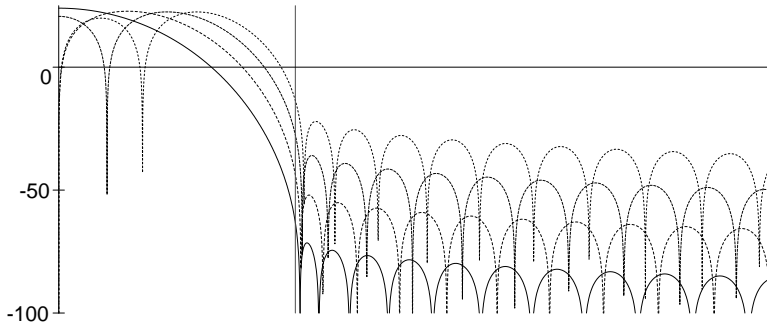


Figure 3. The degree of energy confinement of the first four Slepian tapers of Fig. 2 measured by $10 \log_{10}(W_k^{(N)}(f))$. The thin vertical line corresponds to frequency W . The height of the side-lobes increases with k . The corresponding eigenvalues are $\lambda^{(0)} = 0.9999999997$, $\lambda^{(1)} = 0.9999999693$, $\lambda^{(2)} = 0.9999984555$, $\lambda^{(3)} = 0.9999482420$ (10DP).

sequence is time-limited to $0, 1 \dots N$ there is no way to confine the energy to $(-W, W)$ if $W < \frac{1}{2}$ so in this case $\lambda^{(k)} < 1$. We may label the eigenvalues so that

$$0 < \lambda^{(N-1)} < \dots < \lambda^{(1)} < \lambda^{(0)} < 1. \tag{21}$$

Slepian has shown the remarkable property that the first $2NW\Delta t - 1$

eigenvalues are close to one and so provide excellent energy confinement while thereafter they fall rapidly to zero.

The sequence of Slepian tapers lead to a sequence of spectral estimators

$$\hat{S}_k^{(N)}(f) \equiv \Delta t \left| \sum_{r=1}^N w_r^{(k)} X_r e^{-2\pi i f r \Delta t} \right|^2, \quad (22)$$

where according to the last paragraph we may expect only the first $(2NW\Delta t - 1)$ estimators to be useful.

2.5. Multi-taper methods

As its name implies, multi-taper methods apply multiple tapers to the time series. In this way it is possible to achieve high resolution and accuracy. To be specific, one employs the first K Slepian tapers where $K < (2NW\Delta t - 1)$ so that all the tapers have good energy confinement. The simplest multi-taper spectral estimator is then defined as

$$\hat{S}_{\text{mt}}^{(N)}(f) \equiv \frac{1}{K} \sum_{k=1}^K \hat{S}_k^{(N)}(f). \quad (23)$$

This is equivalent to using a kernel

$$\mathcal{W}_{\text{mt}}^{(N)}(f) = \frac{1}{K} \sum_{k=1}^K \mathcal{W}_k^{(N)}(f). \quad (24)$$

(Alternative definitions are possible, for example, weighting each term inversely by the corresponding eigenvalue but these make little difference to the basic picture and we shall deal solely with this simplest definition; see Percival and Walden for a fuller discussion.)

As the tapers from which the different spectral estimators making up the sum in eq. (23) are orthogonal the terms in the sum are pairwise uncorrelated, that is, they provide *independent* spectral estimates. As a consequence the variance of $\hat{S}_{\text{mt}}^{(N)}(f)$ is smaller than that of each term in the sum by a factor of order $1/K$. In fact, under reasonable assumptions $S_k^{(N)}(f)$ is distributed like $(S(f)/2)\chi_2^2$, while $\hat{S}_{\text{mt}}^{(N)}(f)$ will be distributed like $(S(f)/2K)\chi_{2K}^2$.

The advantage of the multi-taper estimate is seen most clearly with an example. We take a realisation of length $N = 512$ from the AR(4) process

$$X_r = 2.7607X_{r-1} - 3.8106X_{r-2} + 2.6535X_{r-3} - 0.9238X_{r-4} + N_r, \quad (25)$$

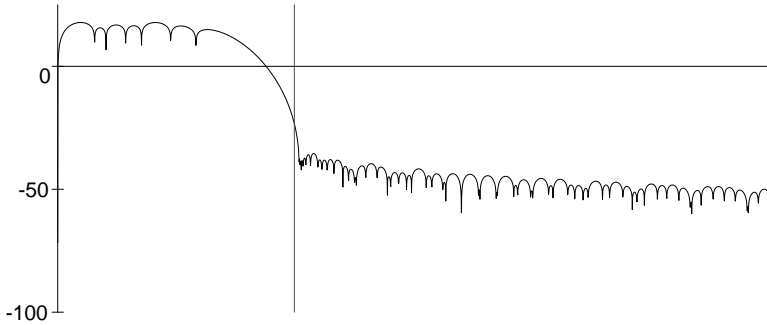


Figure 4. A graph of $10 \log_{10}(\mathcal{W}_{(\text{mt})}^{(N)}(f))$ defined by eq. (24) with $K = 4$. The thin vertical line corresponds to frequency W .

where N_r is Gaussian white noise with zero mean and unit variance. The true spectrum is given by

$$S(f) = 2/|1 - 2.7607e^{2\pi if} + 3.8106e^{4\pi if} - 2.6535e^{6\pi if} + 0.9238e^{8\pi if}|^2. \quad (26)$$

Figure 5 shows a graph of this spectrum together with estimates of it based on a realisation of it of length $N = 512$. Fig. 5(a) shows the severe bias in the periodogram spectral estimate at high frequency due to leakage. Fig. 5(b) shows the marked improvement in bias obtained by windowing the data with a single Slepian taper. Fig. 5(c) shows the marked improvement in the variance in the spectral estimator obtained by multi-tapering.

Multi-taper spectral estimates are implemented in GRASP by the program `multitaper_spectrum()` which was used to generate Figs. 5(b) and (c).

3. HARMONIC ANALYSIS

Multi-taper methods are also extremely well suited to determining the magnitude and location of line features in the spectrum. Suppose for simplicity that our signal consists of a pure harmonic of frequency f_0 plus noise:

$$X(t) = Ae^{i2\pi f_0 t} + N(t)m \quad (27)$$

where A is the complex amplitude of the line. As before we suppose that this signal is sampled at N discrete times at intervals Δt . The obvious

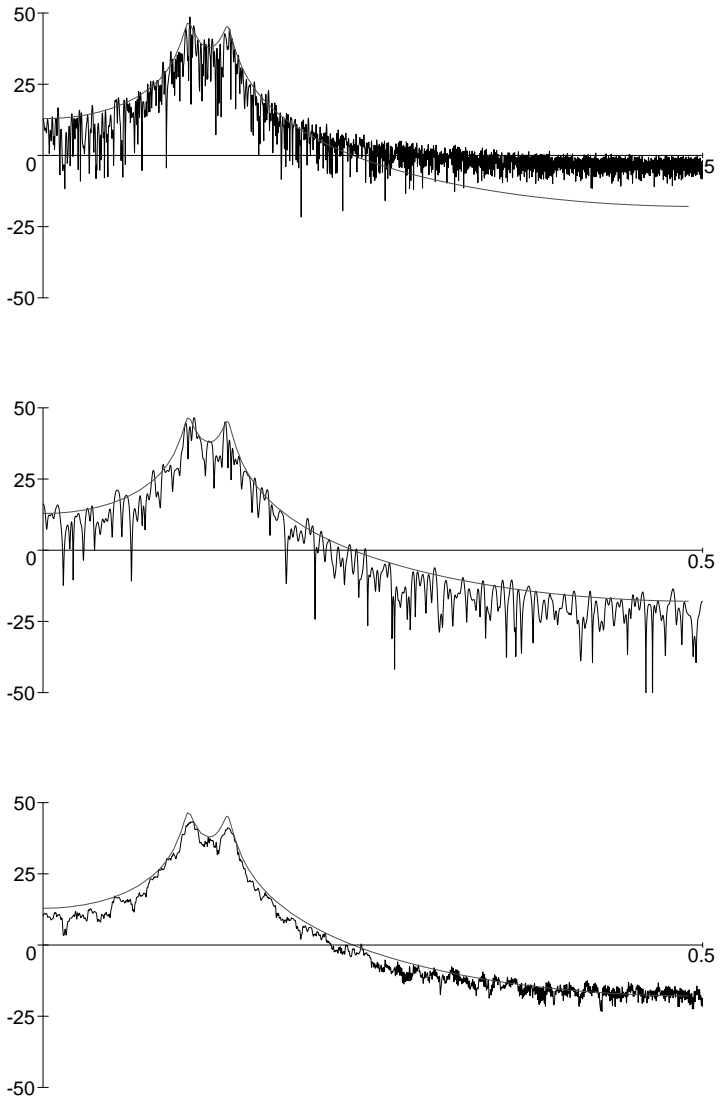


Figure 5. The spectrum of the AR(4) process defined by eq. (25) together with estimates of it based on a realisation of it of length $N = 512$. (a) shows the periodogram (b) the taper spectral estimator $\hat{S}_0^{(512)}(f)$ based on a single Slepian taper, note the improvement in the bias at high frequency, and (c) the multi-taper spectral estimate based on $K = 6$, note the improvement in the variance at high frequency. The latter two were calculated using the GRASP routine `multitaper_spectrum()`.

way to estimate A from the sequence x_1, x_2, \dots, x_N is to minimise the spectral estimator of $X(t) - \hat{A}e^{i2\pi f_0 t}$ at frequency f_0 . Using a windowed periodogram we have

$$\hat{S}_w^{(N)}(f_0) = \Delta t \left| \sum_{r=1}^N w_r (X_r - \hat{A}_w e^{i2\pi f_0 r \Delta t}) e^{-2\pi i f_0 r \Delta t} \right|^2 \tag{28}$$

$$= \Delta t |J(f_0) - \tilde{w}^{(N)}(0) \hat{A}_w|^2, \tag{29}$$

where

$$J(f) \equiv \sum_{r=1}^N w_r X_r e^{-2\pi i f r \Delta t} \quad \text{and} \quad \tilde{w}^{(N)}(0) = \sum_{r=1}^N w_r. \tag{30}$$

This is clearly minimised by taking

$$\hat{A}_w = \frac{J(f_0)}{\tilde{w}^{(N)}(0)}. \tag{31}$$

If we use the multi-taper spectral estimate and minimise

$$\hat{S}_{\text{mt}}^{(N)}(f_0) = \frac{\Delta t}{K} \sum_{k=1}^K |J_k(f_0) - \tilde{w}_k^{(N)}(0) \hat{A}_{\text{mt}}|^2, \tag{32}$$

we obtain

$$\hat{A}_{\text{mt}} = \frac{\sum_{k=1}^K \tilde{w}_k^{(N)}(0) J_k(f_0)}{\sum_{k=1}^K [w_k^{(N)}(0)]^2}, \tag{33}$$

with the obvious notation that $J_k(f_0)$ refers to the quantities defined in eq. (30) for the k th Slepian taper. (Note that, in fact, for k odd the Slepian tapers are odd functions so that $w_k^{(N)}(0) = 0$ for odd k .)

A principal advantage of this method is that the ratio of the variance of the spectrum with the estimated line removed to the estimated power in the line provides a concrete statistical test of whether the assumption of the existence of a line component is justified. To be specific

$$F(f_0) = (K - 1) \frac{|\hat{A}_{\text{mt}}|^2}{\hat{S}_{\text{mt}}^{(N)}(f_0)}, \tag{34}$$

where $S_{\text{mt}}^{(N)}(f_0)$ is given by eq. (32), will have the F -distribution with 2 and $2K - 2$ degrees of freedom. To justify the assumption of a spectral line at frequency f_0 this statistic should exceed, for example, the 99% point of the $F_{2,2K-2}$ distribution.

If the F -test is significant, eq. (32) suggests [7] that we may reshape the spectrum in the region $[f_0 - W, f_0 + W]$ to remove the effect of the line by defining

$$\hat{S}_{\text{mt}}^{(N)}(f) = \frac{\Delta t}{K} \sum_{k=1}^K |J_k(f) - \tilde{w}_k^{(N)}(f - f_0) \hat{A}_{\text{mt}}|^2, \quad (35)$$

$$f_0 - W < f < f_0 + W.$$

Harmonic analysis is implemented within GRASP by the program `remove_spectral_lines()`. This program tests every Fourier frequency for the presence of a line component and returns a list of the frequencies where spectral lines were found together with its amplitude and phase and the corresponding value of the F -test. It also returns the data set with the spectral lines subtracted (if desired) and provides both an ‘initial’ multi-taper spectrum of the original data, and a ‘final’ multi-taper spectrum, after line removal.

As an illustration, in Figure 6 we apply `remove_spectral_lines()` to a realisation of length 512 of the time series

$$X_r = \cos(2\pi(0.2)r) + 0.25 \sin(2\pi(0.35)r) + N_r, \quad (36)$$

where N_r is Gaussian white noise with zero mean and unit variance. Taking $W = 4/(N\Delta t)$ and $K = 6$, `remove_spectral_lines()` estimates the presence of two line features in the data:

```
line of amplitude 0.97 and phase 4.74 at freq 0.200024
(F-test value 136.30)
line of amplitude 0.30 and phase 94.63 at freq 0.350134
(F-test value 26.67)
```

These F -test values correspond to $F_{2,10}$ percentage points of 6×10^{-8} and 1×10^{-4} for the probability of such features in the spectrum on the basis of noise alone. Fig. 6 shows the multi-taper spectrum of the data and the reshaped spectrum after accounting for the presence of these estimated line features.

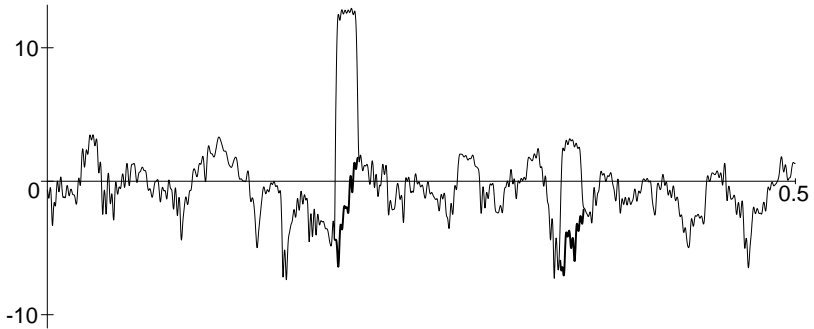


Figure 6. The multi-taper spectral estimate of the time series (36) and the reshaped spectrum (in bold) after the removal of the line features estimated in the data by `remove_spectral_lines()`.

4. CONCLUSION

We have endeavoured to give a simple presentation of multi-taper methods and suggest their usefulness for gravitational wave data analysis problems. For clarity we have presented simple models here to illustrate these methods. The analysis on these models was performed using the multi-taper methods contained the GRASP (Gravitational Radiation Analysis and Simulation Package) software package [1]. For a detailed description of this package and examples of the application of these techniques to data from the Caltech 40m prototype interferometer see the GRASP manual [1].

REFERENCES

1. Allen, B., et al. (1998). GRASP: a data analysis package for gravitational wave detection version 1.8.3. Manual and package at: <http://www.lsc-group.phys.uwm.edu/>.
2. Blackman, R. B., and Tukey, J. W. (1958). *The Measurement of Power Spectra* (Dover, New York).
3. Kay, S. M. (1988). *Modern Spectral Estimation — Theory and Application* (Prentice-Hall, New York).
4. Percival, D. B., and Walden, A. T. (1993). *Spectral Analysis for Physical Applications – Multitaper and Conventional Univariate Techniques* (Cambridge University Press, Cambridge).

5. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1992). *Numerical Recipes — The Art of Scientific Computing* (Cambridge University Press, Cambridge).
6. Slepian, D. (1978). *Bell Syst. Tech. J.* **57**, 1371.
7. Thomson, D. J. (1982). *Proc. IEEE* **70**, 1055.