SPECTRAL ESTIMATION

John Fredy Ochoa Gómez
Bioengineering Department
University of Antioquia

Spectrum estimation using Periodogram, Bartlett and Welch

Guido Schuster

Slides follow closely chapter 8 in the book „Statistical Digital Signal Processing and Modeling“ by Monson H. Hayes and most of the figures and formulas are taken from there

CONTENT

1. **Power Spectral Density**

1. Welch’s method

1. Multitaper estimation

1. Libraries

SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

- The so-called 1/w noises can be defined and characterized by the spectral power.
- The power spectrum describes the distribution of power into frequency components of the signal
- The Power Spectral Density (PSD) refers to the spectral energy distribution that would be found per unit frequency
  - Since the total energy of such a signal over all time would generally be infinite.

SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

- For truth random process, stationary and without finite energy:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \infty$$

- Being stationary, can be characterized by its autocorrelation function

$$\gamma_{xx}(\tau) = E[x(t)x(t+\tau)]$$

- According the Wienner-Khintchine theorem, the PSD is the Fourier transform of the autocorrelation function

$$\Gamma_{xx}(w) = \int_{-\infty}^{\infty} \gamma_{xx}(\tau)\, e^{-j2\pi w\tau} d\tau$$

SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

- Given that we are working with only a realization of the random process we cannot access the real autocorrelation function, so, we need to estimate $\gamma_{xx}(\tau)$ and $\Gamma_{xx}(w)$ using a finite sample of the process x(t).

- During a time interval 2T we can define the the average autocorrelation as:

$$R(\tau) = \frac{1}{2T}\int_0^{2T} x(t)x(t+\tau)dt$$

SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

- Being stationary the process then:

$$\gamma_{xx}(\tau) = \lim_{T \to \infty} R(\tau)$$

- So, the Fourier transform of *R(τ)* is an approximation of the real Power Spectral Density (Periodogram)

$$P(w) = \int_0^{2T} R(\tau) e^{-jw\tau} d\tau = \frac{|X(w)|^2}{2T}$$

- Being the real PSD obtained as:

$$\Gamma_{xx}(w) = \lim_{T \to \infty} E[P(w)]$$

Performance of the Periodogram

- If N goes to infinity, does the Periodogram converge towards the power spectrum in the mean squared sense?

$$\lim_{N\to\infty} E\left\{\left[\hat{P}_{per}(e^{j\omega}) - P_x(e^{j\omega})\right]^2\right\} = 0$$

- Necessary conditions
  - asymptotically unbiased:

$$\lim_{N\to\infty} E\left\{\hat{P}_{per}(e^{j\omega})\right\} = P_x(e^{j\omega})$$

  - variance goes to zero:

$$\lim_{N\to\infty} \mathrm{Var}\left\{\hat{P}_{per}(e^{j\omega})\right\} = 0$$

- In other words, it must be a consistent estimate of the power spectrum

SPECTRAL ESTIMATION

Periodogram bias

To compute the bias we first find the expected value of the autocorrelation estimate

$$E\left\{\hat{r}_x(k)\right\} = \frac{1}{N}\sum_{n=0}^{N-1-k} E\left\{x(n+k)x^*(n)\right\}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1-k} r_x(k) = \frac{N-k}{N} r_x(k)$$

Hence the estimate of the autocorrelation is biased with a triangular window (Bartlett)

$$E\left\{\hat{r}_x(k)\right\} = w_B(k) r_x(k)$$

$$w_B(k) = \begin{cases} \dfrac{N-|k|}{N} & ; \quad |k| \le N \\ 0 & ; \quad |k| > N \end{cases}$$

SPECTRAL ESTIMATION

Periodogram bias

The expected value of the Periodogram can now be calculated:

$$E\left\{\hat{P}_{per}(e^{j\omega})\right\} = E\left\{\sum_{k=-N+1}^{N-1} \hat{r}_x(k)e^{-jk\omega}\right\}$$

$$= \sum_{k=-N+1}^{N-1} E\left\{\hat{r}_x(k)\right\} e^{-jk\omega}$$

$$= \sum_{k=-\infty}^{\infty} r_x(k) w_B(k) e^{-jk\omega}$$

Thus the expected value of the Periodogram is the convolution of the power spectrum with the Fourier transform of a Bartlett window

$$\boxed{E\left\{\hat{P}_{per}(e^{j\omega})\right\} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega})}$$

$$W_B(e^{j\omega}) = \frac{1}{N}\left[\frac{\sin(N\omega/2)}{\sin(\omega/2)}\right]^2$$

SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

$$\hat{P}_{per}(e^{j\omega}) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-jn\omega}\right|^2$$

Bias

$$E\left\{\hat{P}_{per}(e^{j\omega})\right\} = \frac{1}{2\pi} P_x(e^{j\omega}) * W_B(e^{j\omega})$$

Resolution

$$\Delta\omega = 0.89\frac{2\pi}{N}$$

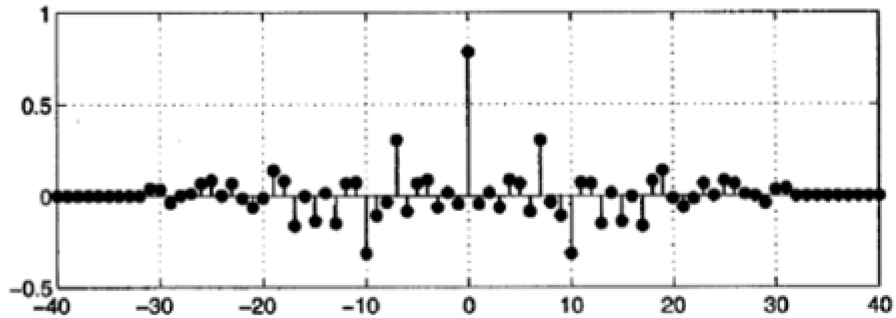
Variance

$$\mathrm{Var}\left\{\hat{P}_{per}(e^{j\omega})\right\} \approx P_x^2(e^{j\omega})$$

SPECTRAL ESTIMATION

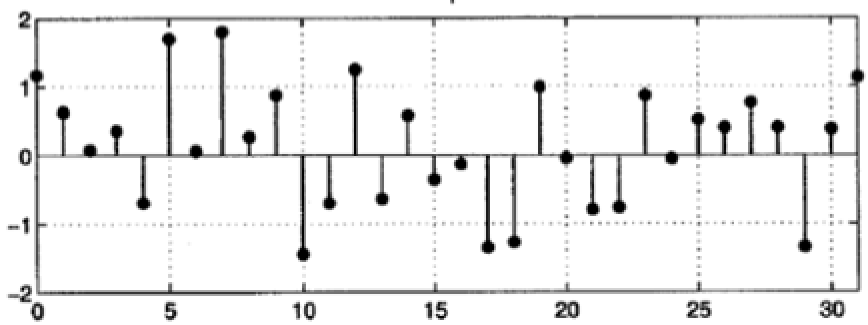
POWER SPECTRAL DENSITY

Periodogram of white noise

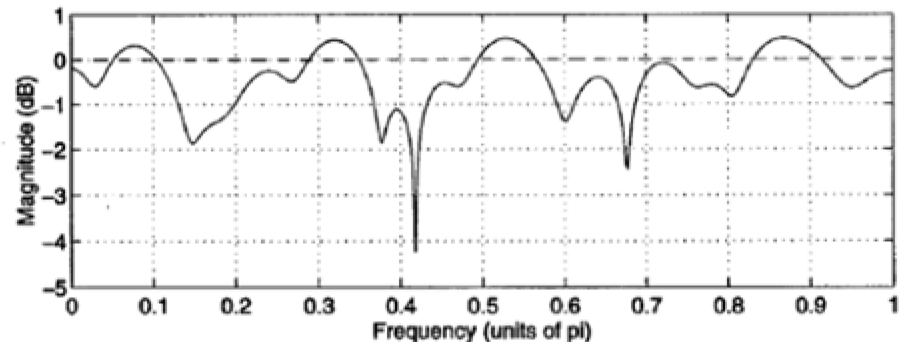


$$r_x(k) = \sigma_x^2 \delta(k)$$

32 samples



$$P_x(e^{j\omega}) = \sigma_x^2$$



Example: Periodogram of a Sinusoidal in Noise

- Consider a a random process consisting of a sinusoidal in white noise, where the phase of the sinusoidal is uniformly $[-\pi, \pi]$ distributed and $A=5$, $\omega_0=0.4\pi$

- N=64 on top
  and N=256 on the bottom

- Overlay of 50 Periodogram on the left and average on the right

$$x(n) = A\sin(n\omega_0 + \phi) + v(n)$$

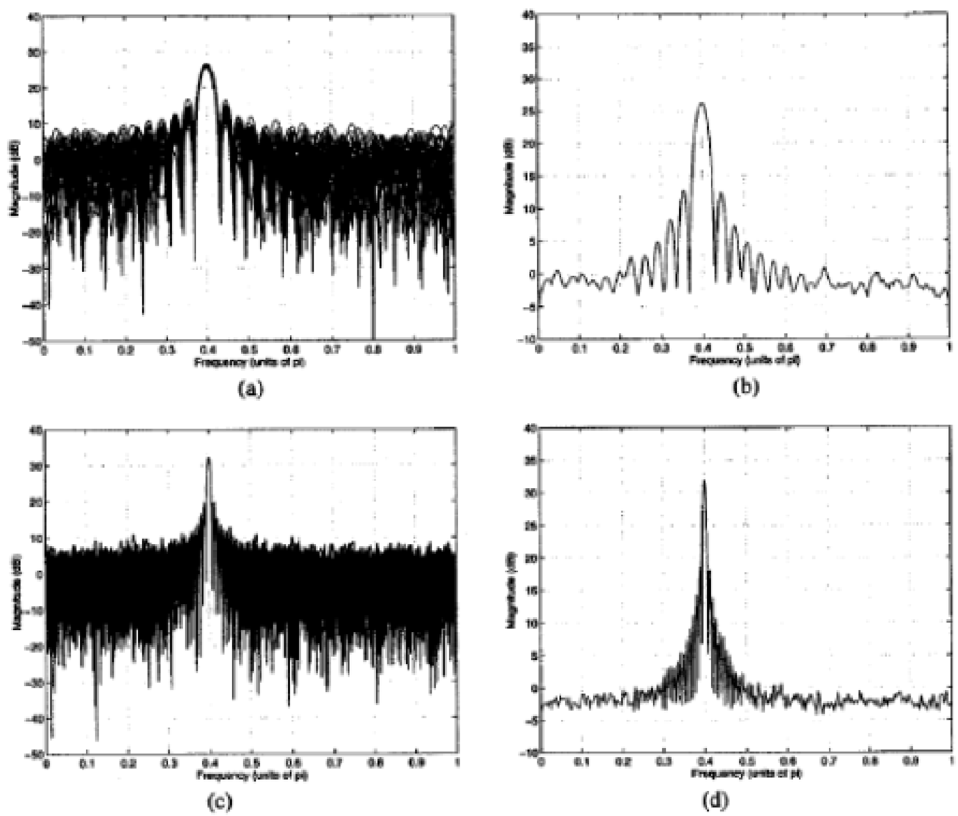


Figure 8.6 *The periodogram of a sinusoid in white noise. (a) Overlay plot of 50 periodograms using $N = 64$ data values and (b) the periodogram average. (c) Overlay plot of 50 periodograms using $N = 256$ data values and (d) the periodogram average.*

Example: Periodogram of two Sinusoidal in Noise

- Consider a a random process consisting of two sinusoidal in white noise, where the phases of the sinusoidal are uniformly [$-\pi$, π] distributed and
  A=5, $\omega_1=0.4\pi$, $\omega_2=0.45\pi$

- N=40 on top
  and N=64 on the bottom

- Overlay of 50 Periodogram on the left and average on the right

$$x(n) = A\sin(n\omega_1 + \phi_1) + A\sin(n\omega_2 + \phi_2) + v(n)$$

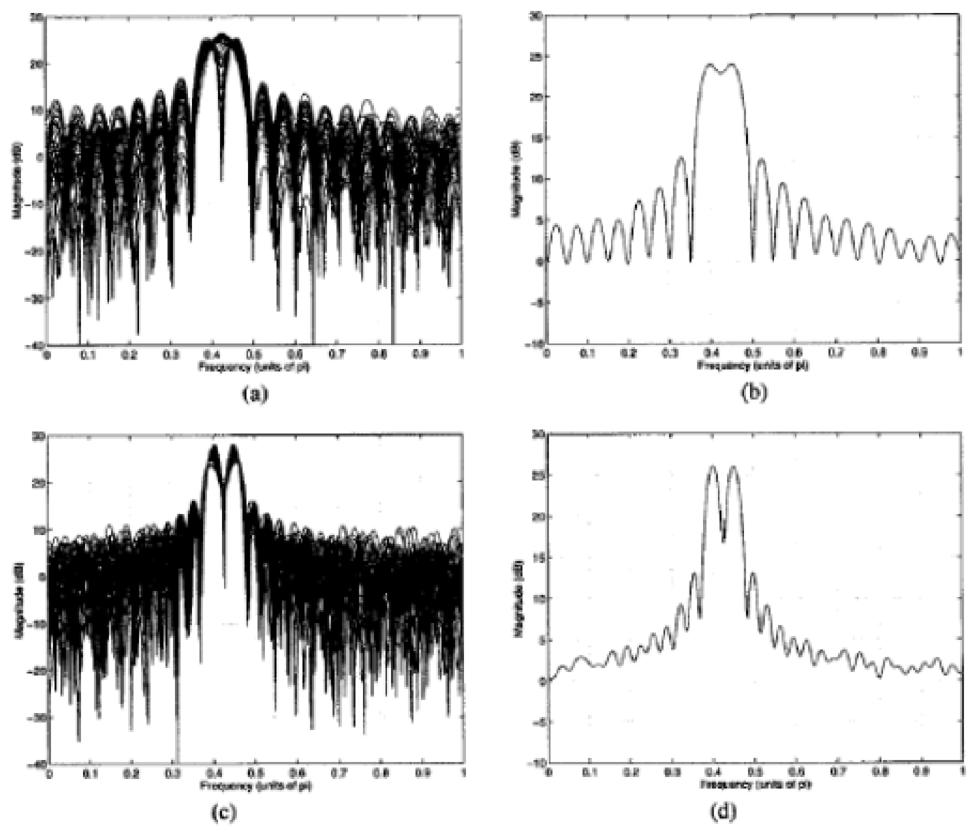


Figure 8.8 *The periodogram of two sinusoids in white noise with $\omega_1 = 0.4\pi$ and $\omega_2 = 0.45\pi$. (a) Overlay plot of 50 periodograms using $N = 40$ data values and (b) the ensemble average. (c) Overlay plot of 50 periodograms using $N = 64$ data values and (d) the ensemble average.*

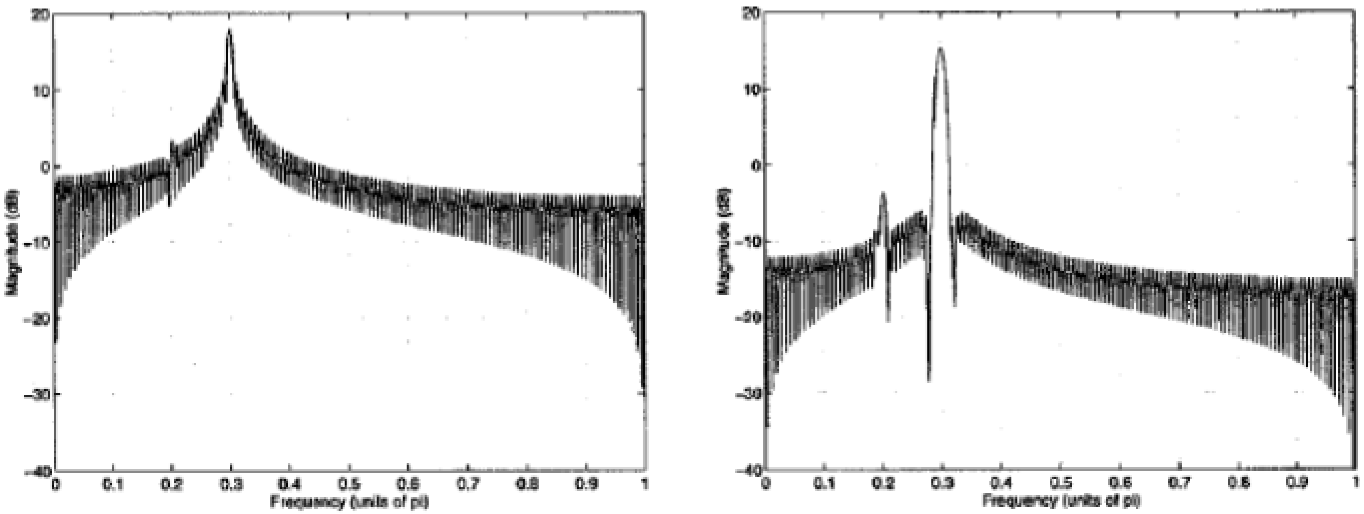
The modified Periodogram

Smoothing is determined by the window that is applied to the data

While the rectangular window as the smallest main lobe of all windows, its sidelobes fall off rather slowly

$$x(n) = 0.1\sin(n\omega_1 + \phi_1) + \sin(n\omega_2 + \phi_2) + v(n)$$

Rectangular data window   Hamming data window



(a)   (b)

Figure 8.10 *Spectral analysis of two sinusoids in white noise with sinusoidal frequencies of $\omega_1 = .2\pi$ and $\omega_2 = .3\pi$ and a data record length of $N = 128$ points. (a) The expected value of the periodogram. (b) The expected value of the modified periodogram using a Hamming data window.*

The modified Periodogram

- Nothing is free. As you notice, the Hamming window has a wider main lobe

- The Periodogram of a process that is windowed with a general window is called modified Periodogram

$$\boxed{\hat{P}_M(e^{j\omega}) = \frac{1}{NU}\left|\sum_{n=-\infty}^{\infty} x(n)w(n)e^{-jn\omega}\right|^2}$$

- N is the length of the window and U is a constant that is needed so that the modified Periodogram is asymptotically unbiased

$$U = \frac{1}{N}\sum_{n=0}^{N-1} |w(n)|^2$$

Modified periodogram summary

Table 8.3 Properties of the Modified Periodogram

	$$\hat{P}_M(e^{j\omega}) = \frac{1}{NU}\left|\sum_{n=-\infty}^{\infty} w(n)x(n)e^{-jn\omega}\right|^2$$
	$$U = \frac{1}{N}\sum_{n=0}^{N-1}|w(n)|^2$$
Bias	$$E\left\{\hat{P}_M(e^{j\omega})\right\} = \frac{1}{2\pi NU}P_x(e^{j\omega}) * |W(e^{j\omega})|^2$$
Resolution	Window dependent
Variance	$$\text{Var}\left\{\hat{P}_M(e^{j\omega})\right\} \approx P_x^2(e^{j\omega})$$

Bartlett's method

- Averaging (sampe mean) a set of uncorrelated measurements of a random variable results in a consistent estimate of its mean
- In other words: Variance of the sample mean is inversely proportional to the number of measurements
- Hence this should also work here, by averaging Periodograms

This suggests that we consider estimating the power spectrum of a random process by periodogram averaging. Thus, let $x_i(n)$ for $i = 1, 2, \ldots, K$ be K uncorrelated realizations of a random process $x(n)$ over the interval $0 \le n < L$. With $\hat{P}_{per}^{(i)}(e^{j\omega})$ the periodogram of $x_i(n)$,

$$\hat{P}_{per}^{(i)}(e^{j\omega}) = \frac{1}{L}\left|\sum_{n=0}^{L-1} x_i(n)e^{-jn\omega}\right|^2 \quad ; \quad i = 1, 2, \ldots, K$$

Bartlett's method

- Now we reap the reward: the variance is going to zero as the number of subsequences goes to infinity
- If both, K and L go to infinity, this will be a consistent estimate of the power spectrum
- In addition, for a given N=K*L, we can trade off between good spectral resolution (large L) and reduction in variance (Large K)

$$\mathrm{Var}\left\{\hat{P}_B(e^{j\omega})\right\} \approx \frac{1}{K}\mathrm{Var}\left\{\hat{P}_{per}^{(i)}(e^{j\omega})\right\} \approx \frac{1}{K}P_x^2(e^{j\omega})$$

Table 8.4 *Properties of Bartlett's Method*

$$\hat{P}_B(e^{j\omega}) = \frac{1}{N}\sum_{i=0}^{K-1}\left|\sum_{n=0}^{L-1} x(n+iL)e^{-jn\omega}\right|^2$$

Bias

$$E\left\{\hat{P}_B(e^{j\omega})\right\} = \frac{1}{2\pi}P_x(e^{j\omega}) * W_B(e^{j\omega})$$

Resolution

$$\Delta\omega = 0.89K\frac{2\pi}{N}$$

Variance

$$\mathrm{Var}\left\{\hat{P}_B(e^{j\omega})\right\} \approx \frac{1}{K}P_x^2(e^{j\omega})$$

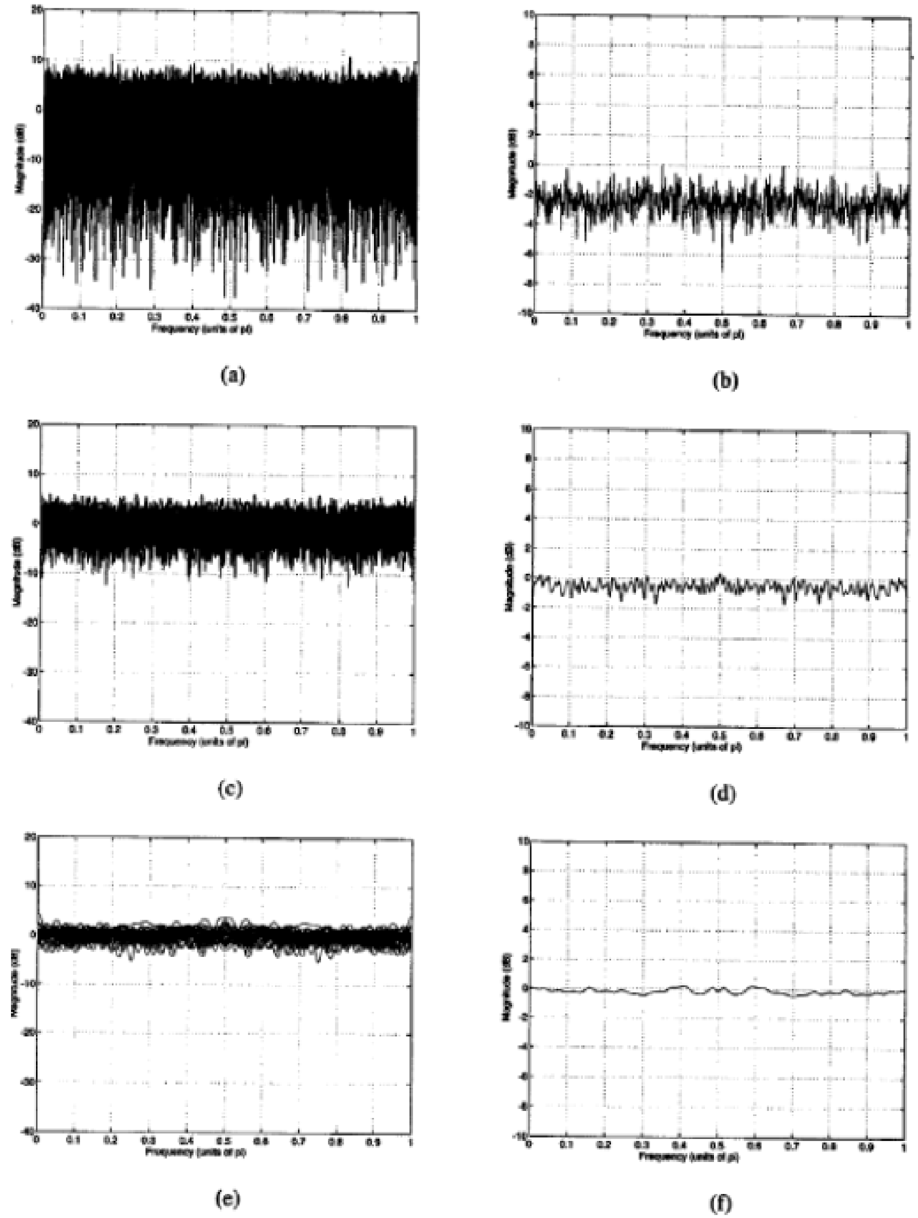
Bartlett's method: White noise

- a) Periodogram with N=512
- b) Ensemble average

- c) Overlay of 50 Bartlett estimates with K=4 and L=128
- d) Ensemble average

- e) Overlay of 50 Bartlett estimates with K=8 and L=64
- f) Ensemble average

Overlay of 50 estimates Ensemble Average



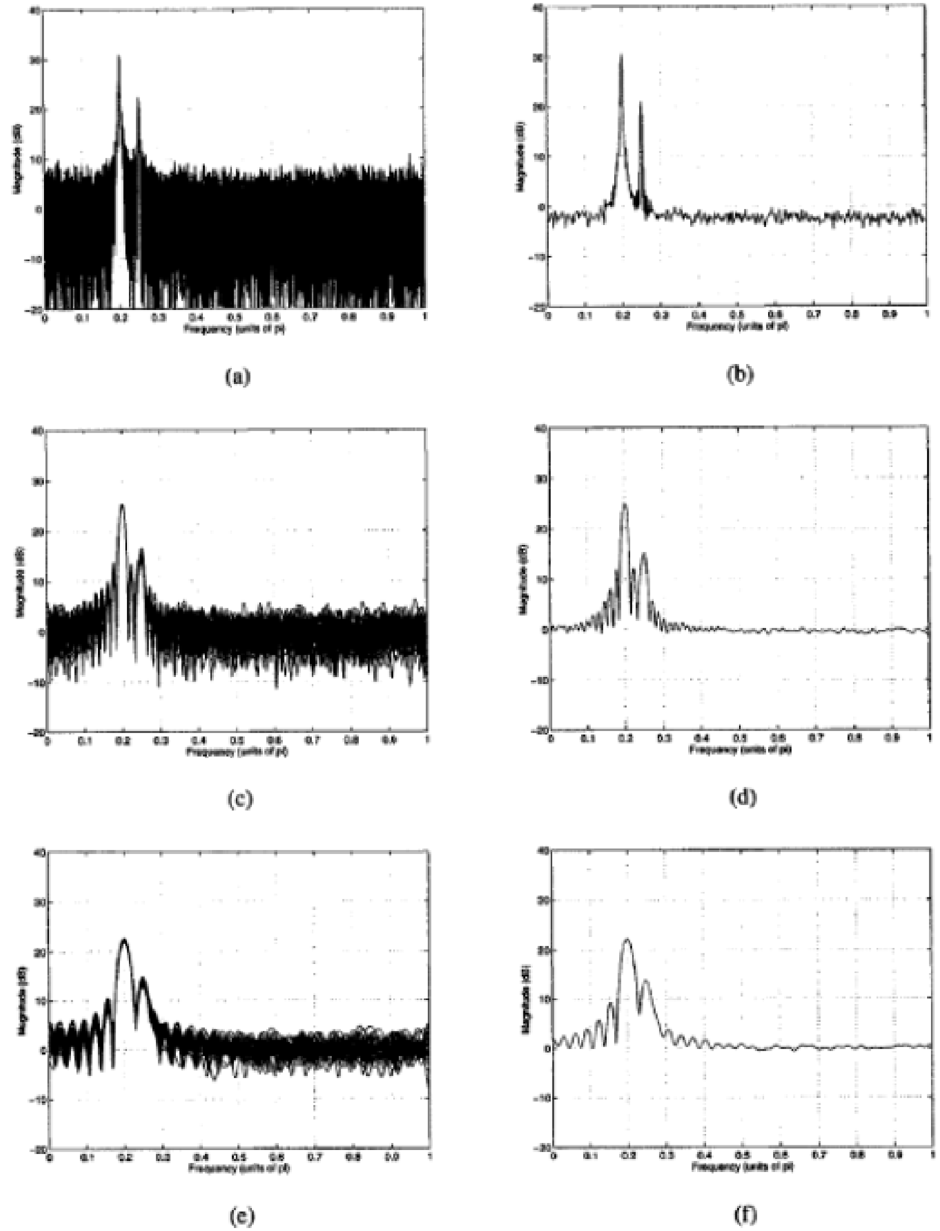
Bartlett's method: Two sinusoidal in white noise

- a) Periodogram with N=512
- b) Ensemble average

- c) Overlay of 50 Bartlett estimates with K=4 and L=128
- d) Ensemble average

- e) Overlay of 50 Bartlett estimates with K=8 and L=64
- f) Ensemble average

Note how larger K results in shorter L and hence in less spectral resolution



SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

- This spectral estimate has two issues:

1. the bias, which implies that the estimate is not equal to the real spectrum;

1. the variance, which implies that the estimate would not converge to the true estimate even increasing the amount of data.

SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

- The bias comes in two forms:

1. narrowband bias, where the error is related to interaction of nearby frequencies;

1. and broadband bias, where the error is due to interaction between distant frequencies.

SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

- Although the bias problem diminishes as $T \rightarrow \infty$, in practice the effect of increasing T is limited by the lack of stationarity in the biosignals.
- Bias can be reduced by multiplying the data by a taper (window) function, $w[n]$, before transforming the series to the frequency domain.
- Variance is usually addressed by averaging overlapping segments of the time series.

SPECTRAL ESTIMATION

POWER SPECTRAL DENSITY

- Using tapers functions, the influence of distant frequencies can be reduced at the expense of blurring the spectrum over nearby frequencies.

- The practice is justified under the assumption that the true spectrum is locally constant.

CONTENT

1. Power Spectral Density

1. **Welch's method**

1. Multitaper estimation

1. Libraries

SPECTRAL ESTIMATION

WELCH'S METHOD

- Welch's periodogram is carried out by dividing the signal into successive blocks, forming the periodogram for each block, and averaging.

- For a segment m of M samples, that is a windowed version of the original signal, we have:

$$x_m[n] := w[n]x[n+mR], \qquad n = 0,1,\dots,M-1;\ m = 0,1,\dots,Q-1$$

  - Being R the hop size (more usually used the parameter overlap = M – R)
  - Q the number of segments

- The Periodogram is defined as: $P_i[k] = \frac{1}{M}|X[k]|^2$

SPECTRAL ESTIMATION

WELCH'S METHOD

- And the Welch's Periodogram is estimated as the average of the Periodogram obtained from Q segments:

$$S_{xx}[k] = \frac{1}{Q}\sum_{q=1}^{Q} P_q[k]$$

- For a signal with a total number of points N_x we have a total number of segments:

$$Q = \left\lfloor \frac{N_x - M}{R} \right\rfloor + 1$$

- Where *x* stands for the largest integer less than or equal to x

SPECTRAL ESTIMATION

WELCH'S METHOD

- The window function w[n] is used to reduce the side-lobe level in the spectral density estimate, at the expense of frequency resolution (because the reduction of samples used).

  - When using a non-rectangular analysis window, *R* cannot exceed half the frame length (M/2) without introducing a non-uniform sensitivity to the data x[n] over time.
  - For the rectangular window, we can set *R=M* and have non-overlapping windows.
  - For Hamming, Hanning, and any other generalized Hamming normally is used *R=(M-1)/2* for odd-length windows.
  - For the Blackmann window, *R≈M/3* is typical.

Welch's method summary

Table 8.5 Properties of Welch's Method

$$\hat{P}_W(e^{j\omega}) = \frac{1}{KLU}\sum_{i=0}^{K-1}\left|\sum_{n=0}^{L-1} w(n)x(n+iD)e^{-jn\omega}\right|^2$$

$$U = \frac{1}{L}\sum_{n=0}^{L-1}|w(n)|^2$$

Bias

$$E\left\{\hat{P}_W(e^{j\omega})\right\} = \frac{1}{2\pi LU}P_x(e^{j\omega}) * |W(e^{j\omega})|^2$$

Resolution Window dependent

Variance†

$$\mathrm{Var}\left\{\hat{P}_W(e^{j\omega})\right\} \approx \frac{9}{16}\frac{L}{N}P_x^2(e^{j\omega})$$

† Assuming 50% overlap and a Bartlett window.

CONTENT

1. Power Spectral Density

1. Welch's method

1. **Multitaper estimation**

1. Libraries

SPECTRAL ESTIMATION

MULTITAPER ESTIMATION

- The multitaper spectrum tries to solve the bias and variance problem using several orthogonal tapers.

- The simplest example of the method is given by the direct multitaper estimate, $S_{MT}(f)$, defined as the average of K individual tapered spectral estimates:

$$S_{MT}(f) = \frac{1}{K}\sum_{k=1}^{K} |X_k(f)|^2$$

$$X_k(f) = \sum_{t=0}^{N-1} w_t(k) x_t e^{-2\pi i f k \Delta}$$

- The $w_t(k)$ (k = 1, 2, …, K) constitutes K orthogonal taper functions.

SPECTRAL ESTIMATION

MULTITAPER ESTIMATION

- A particular choice for these taper functions, with optimal **spectral concentration properties**, is given by the discrete prolate spheroidal sequences or "Slepian functions".

- Let $w_t(k, W, N)$ be the kth Slepian function of length N and frequency bandwidth parameter W.

- The Slepians form an orthogonal basis set for sequences of length, N, and be characterized by a bandwidth parameter W.

SPECTRAL ESTIMATION

MULTITAPER ESTIMATION

- For a given bandwidth parameter W and taper length N,

  $K = 2NW - 1$

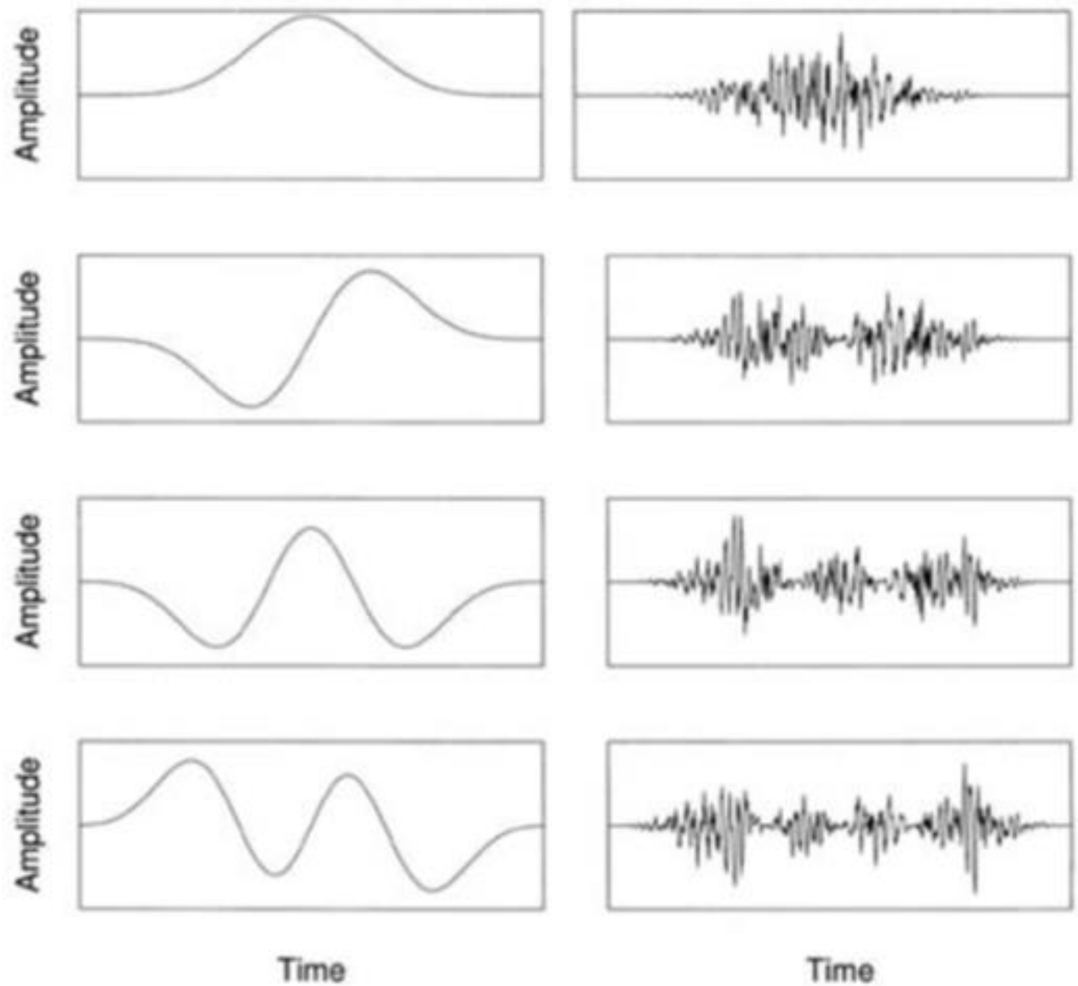
  sequences are obtained each having their energy effectively concentrated within a range [–W, W] of frequency.

- The remaining functions have progressively worsening spectral concentration properties.

SPECTRAL ESTIMATION

MULTITAPER ESTIMATION

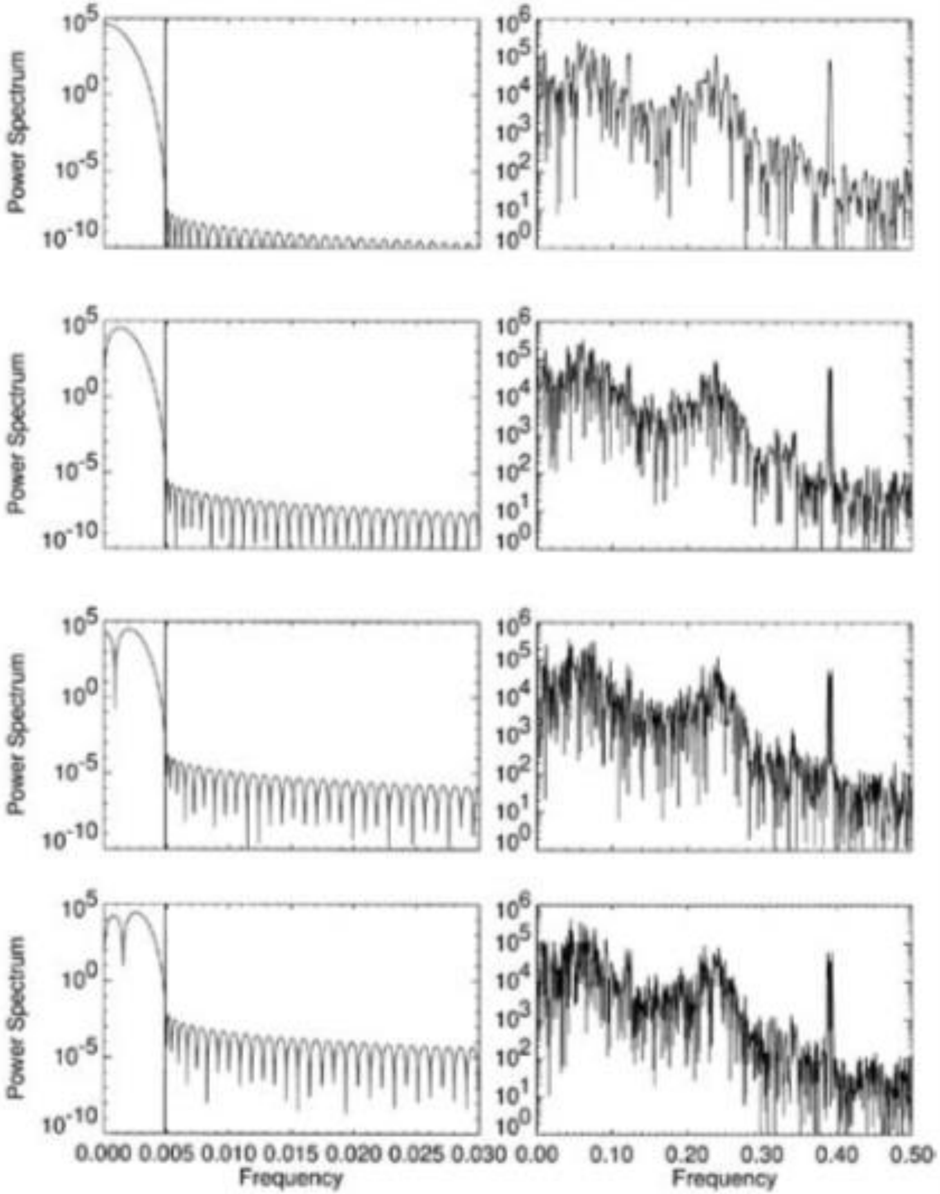
- First four Slepian functions for $W = 5/N$ (left) and the application of the taper to the data (right)



SPECTRAL ESTIMATION

MULTITAPER ESTIMATION

- Spectra of the data tapers, displaying the spectral concentration property.
- The vertical marker denotes the bandwidth parameter *W*.
- The right column display the magnitude-square of the FFT



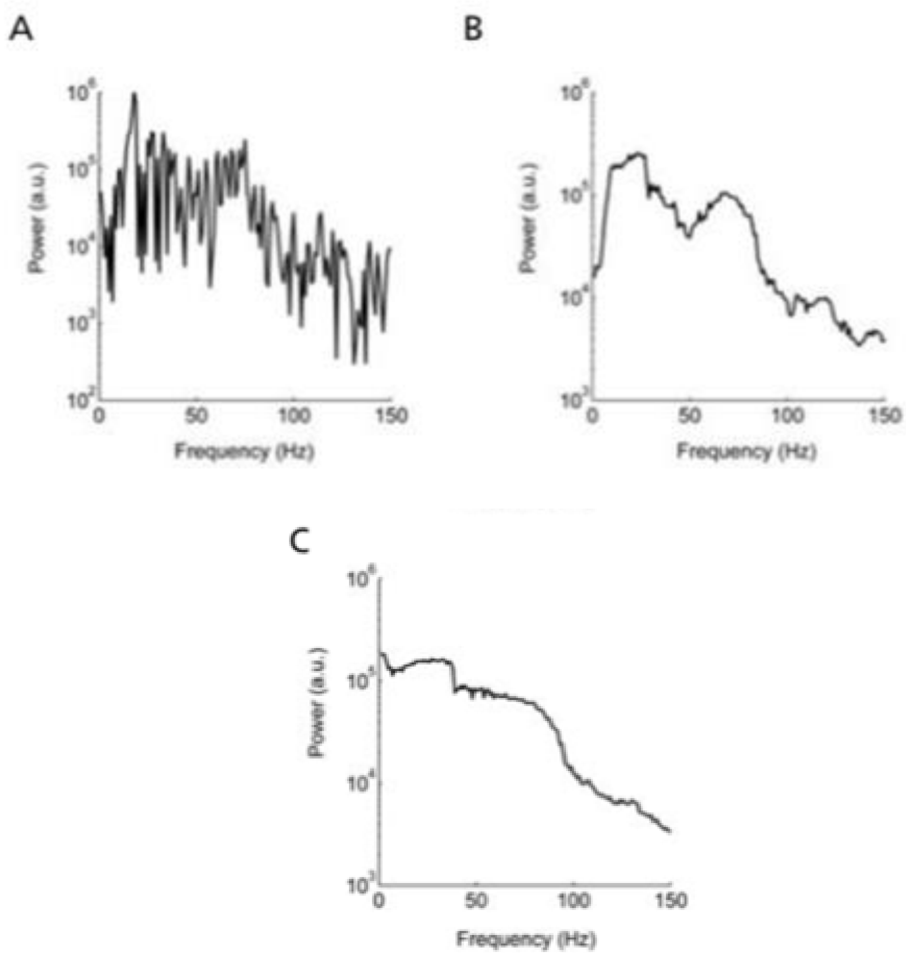
SPECTRAL ESTIMATION

MULTITAPER ESTIMATION

(A) Periodogram estimate of the spectrum based on a single trial of LFP activity. The variability in the estimate is significant.

(B) Multitaper estimate of the spectrum on the same data with *W* = 10 Hz, averaged across 9 tapers. This estimate is much smoother and reveals the presence of two broad peaks at 20 Hz and 60 Hz.

(C) Multitaper spectrum estimate with *W* = 20 Hz. This estimate is even smoother, which reflects the increased number of tapers available to average across (19 instead of 9). However, the assumption that the spectrum is constant with the 20 Hz bandwidth is clearly wrong and leads to noticeable distortion in the spectrum (narrow-band bias).



SPECTRAL ESTIMATION

MULTITAPER ESTIMATION

- The choice of the time window length *N* and the bandwidth parameter *W* is critical for applications.

- Depend on the data set at hand, and are best made iteratively using visual inspection and some degree of trial and error.

- The variance in the estimate is typically reduced by $2NW$.

SPECTRAL ESTIMATION

MULTITAPER ESTIMATION

- The choice of *W* is a choice of how much to smooth.
  - The bandwidth parameter should be chosen to reduce variance while not overly distorting the spectrum by increasing narrow-band bias.

- As a rule, a reasonable strategy is fix the time bandwidth product *NW* at a small number (typically 3 or 4), and then vary the window length in time until sufficient spectral resolution is obtained.

- The best choice of time-frequency resolution will depend on
  - the frequency band of interest
  - the temporal dynamics in the signal

EXAMPLES

WELCH METHOD USING SCIPY

scipy.signal.welch

scipy.signal.welch(*x, fs=1.0, window='hanning', nperseg=256, noverlap=None, nfft=None, detrend='constant', return_onesided=True, scaling='density', axis=-1*) [source]

Estimate power spectral density using Welch's method.

Welch's method [R145] computes an estimate of the power spectral density by dividing the data into overlapping segments, computing a modified periodogram for each segment and averaging the periodograms.

Parameters:

x : *array_like*
Time series of measurement values

fs : *float, optional*
Sampling frequency of the *x* time series in units of Hz. Defaults to 1.0.

window : *str or tuple or array_like, optional*
Desired window to use. See get_window for a list of windows and required parameters. If *window* is array_like it will be used directly as the window and its length will be used for nperseg. Defaults to 'hanning'.

nperseg : *int, optional*
Length of each segment. Defaults to 256.

noverlap: int, optional
Number of points to overlap between segments. If None, `noverlap = nperseg / 2`. Defaults to None.

nfft : *int, optional*
Length of the FFT used, if a zero padded FFT is desired. If None, the FFT length is *nperseg*. Defaults to None.

detrend : *str or function, optional*
Specifies how to detrend each segment. If detrend is a string, it is passed as the `type` argument to detrend. If it is a function, it takes a segment and returns a detrended segment. Defaults to 'constant'.

return_onesided : *bool, optional*
If True, return a one-sided spectrum for real data. If False return a two-sided spectrum. Note that for complex data, a two-sided spectrum is always returned.

scaling : *{ 'density', 'spectrum' }, optional*
Selects between computing the power spectral density ('density') where Pxx has units of V**2/Hz if x is measured in V and computing the power spectrum ('spectrum') where Pxx has units of V**2 if x is measured in V. Defaults to 'density'.

axis : *int, optional*
Axis along which the periodogram is computed; the default is over the last axis (i.e. `axis=-1`).

EXAMPLES

MULTITAPER SPECTRA USING SPECTRUM

pmtm(*x, NW=None, k=None, NFFT=None, e=None, v=None, method='adapt', show=False*) [source]

Multitapering spectral estimation

Parameters:

- **x** (*array*) – the data
- **NW** (*float*) – The time half bandwidth parameter (typical values are 2.5,3,3.5,4). Must be provided otherwise the tapering windows and eigen values (outputs of dpss) must be provided
- **k** (*int*) – uses the first k Slepian sequences. If *k* is not provided, *k* is set to *NW**2.
- **NW** –
- **e** – the window concentrations (eigenvalues)
- **v** – the matrix containing the tapering windows
- **method** (*str*) – set how the eigenvalues are used. Must be in ['unity', 'adapt', 'eigen']
- **show** (*bool*) – plot results

Returns: Sk (complex), weights, eigenvalues

```
from spectrum import data_cosine, dpss, pmtm

data = data_cosine(N=2048, A=0.1, sampling=1024, freq=200)
# If you already have the DPSS windows
[tapers, eigen] = dpss(2048, 2.5, 4)
res = pmtm(data, e=eigen, v=tapers, show=False)
# You do not need to compute the DPSS before end
res = pmtm(data, NW=2.5, show=False)
res = pmtm(data, NW=2.5, k=4, show=True)
```