4F7 Spectrum Estimation

Improving the Spectral Estimate

Sumeetpal Singh

Email: sss40@eng.cam.ac.uk

Modification history:
≤2008: S. Godsill, P. Rayner
>2008: S.S. Singh
1 Improving the Spectral Estimate

- The periodogram is a useful tool, but its variability is very high.

- We will consider several common methods to improve the performance, based on averaging, smoothing, and windowing.
2 The Bartlett Procedure

- When the power spectrum was introduced, it was mentioned that (Hayes, pg 99)

\[
S_X(e^{j\omega T}) = \lim_{N \to \infty} \frac{1}{2N + 1} E \left\{ \left| \sum_{n=-N}^{N} x_n e^{-jn\omega T} \right|^2 \right\}. \tag{1}
\]

- It would seem natural to try and improve the spectrum estimate by performing some averaging in order to mimic the ensemble average above. Note that periodogram implements (1) but without the expectation operator.

- Let the data sequence \( x_n \) be of length \( N_s = KN \) and segment this sequence into \( K \) subsequences of length \( N \):

\[
x_n^{(k)} = x_{n+kN}, \quad 0 \leq n \leq N-1, \quad 0 \leq k \leq K-1
\]
• Calculate the periodogram for each frame, denoted by \( \hat{S}_X^{(k)}(e^{j\omega T}) \), \( k = 0, 1, 2, \ldots, K - 1 \).

• The Bartlett estimate is then given by:

\[
\hat{S}_X^B(e^{j\omega T}) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_X^{(k)}(e^{j\omega T})
\]  

(2)

• Figure 1 shows the results from applying the Bartlett procedure to spectrum analysis of a white noise sequence with subsequence length \( N = 128 \) with averaging over 1, 10 and 100 subsequence spectra.

• If the data subsequences are uncorrelated with one another the Bartlett procedure reduces the variance by a factor of \( K \), by less if they are correlated.

• Recall that

\[
E[\hat{S}_X(e^{j\omega T})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{j\theta})S_X(e^{j(\omega T-\theta)}) \, d\theta
\]

(3)
\[ W(e^{j\omega T}) = \frac{1}{N} \left[ \frac{\sin \left( \frac{N\omega T}{2} \right)}{\sin \left( \frac{\omega T}{2} \right)} \right]^2 \xrightarrow{N \to \infty} 2\pi\delta(\omega T) \]

- Bartlett allows a trade-off between frequency resolution (\( \propto N \)) and variance of the estimate (\( \propto 1/K \)).

- Reduction in variance is at the expense of requiring more data for the same resolution.
Figure 1: Bartlett Procedure applied to white noise
3 The Welch Procedure

- The Welch procedure performs averaging over frames as in the Bartlett method

- However, the periodograms are modified to incorporate a window function on the data:

\[
\hat{S}'(k)(e^{j\omega T}) = \frac{1}{N} \left| \sum_{n=0}^{N-1} w_n x_n^{(k)} e^{-j\omega nT} \right|^2
\]

subject to the constraint \(1/N \sum_{n=0}^{N-1} w_n^2 = 1\) to ensure asymptotic unbiasedness

- As for the Bartlett method, averaging is then performed over \(K\) frames:

\[
\hat{S}^W_X(e^{j\omega T}) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}'(k)(e^{j\omega T})
\] (4)
• Also, the frames can overlap. This means \( N_s = KN \) (\( K, N \) defined in the Bartlett method) data points can be used to construct \( K \) frames with \( N' \) data points each where \( N' > N \) or \( N_s = K'N \) where \( K' > K \).

• The first choice can still reduce the variance while the longer frames improves resolution. The second choice retains the resolution of the Bartlett method but with even less variance.

• For a fixed amount of data \( N_s \), frame length \( N \) and 50\% overlap (Hayes, page 418)

\[
\text{var}(\hat{S}_X^W(e^{j\omega T})) = \frac{9N}{16N_s}S_X(e^{j\omega T})^2
\]

• If \( w_n = 1, \ 0 \leq n < N \), and \( w_n = 0 \) outside this interval, the periodogram is obtained.
• To appreciate the effects of different windows, consider the expected value of the estimate.

• The expected value of this spectral estimate can be shown to be (see next page):

\[
E[\hat{S}_X^W(e^{j\omega T})] = \frac{1}{N} \frac{1}{2\pi} V(e^{j\omega T}) \ast S_X(e^{j\omega T})
\]

where \( W(e^{j\omega T}) \) is the DTFT of the window and \( V(e^{j\omega T}) = |W(e^{j\omega T})|^2 \).

• When the segments are non-overlapping the variance is approximately that of the Bartlett estimate.

• When \( w_n \) is the rectangular window (which corresponds to the periodogram),

\[
|W(e^{j\omega T})|^2 = \frac{\sin^2 \left( \frac{N\omega T}{2} \right)}{\sin^2 \left( \frac{\omega T}{2} \right)}
\]
• The window will trade-off spectral resolution (main lobe width) and spectral masking (side lobe amplitude)
To compute $E[\hat{S}_X^W(e^{j\omega T})]$, first expand

$$\left| \sum_{n=0}^{N-1} w_n x_n^{(k)} e^{-j\omega nT} \right|^2$$

(ignoring superscript $(k)$)

$$\frac{1}{N} E \left\{ \left( \sum_{n=-\infty}^{\infty} w_n x_n e^{-j\omega nT} \right) \left( \sum_{m=-\infty}^{\infty} w_m x_m e^{-j\omega mT} \right)^* \right\}$$

$$= \frac{1}{N} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} w_n w_m E \{ x_n x_m \} e^{-j\omega T(n-m)}$$

$$= \frac{1}{N} \sum_{k=-\infty}^{\infty} \left[ \sum_{n=-\infty}^{\infty} w_n w_{n-k} R_{XX}[n-m] \right] v_k$$

$$= \frac{1}{N 2\pi} S_X(e^{j\omega T}) * V(e^{j\omega T})$$

As $v_k = (\{w_n\} * \{w_{-n}\})(k)$, $V(e^{j\omega T}) = |W(e^{j\omega T})|^2$
4 The Blackman-Tukey (BT) Procedure

- Recall from a previous lecture that
  $$\hat{R}_{XX}[l] = \frac{1}{N} \sum_{n=0}^{N-1-l} x_n x_{n+l}$$
  (or $\frac{1}{N-l}$ for the unbiased estimate)

- Given $N$, $\hat{R}_{XX}[l]$, $l = 0, \ldots, N - 1$, is estimated and then the periodogram estimate $\hat{S}_X$ is computed.

- The Blackman-Tukey method reduces the variance of $\hat{S}_X$ by decreasing the contribution estimates of $\hat{R}_{XX}[l]$ for large $l$ make to $\hat{S}_X$

- The Blackman-Tukey method applies a window function of length $2L+1$ to the estimated autocorrelation
function:

$$\hat{S}_X^{BT}(e^{j\omega T}) = \sum_{l=-L}^{L} w_l \hat{R}_{XX}[l] \exp(-jl\omega T)$$

(5)

where $L < N$ and $w_l$ is any suitable window function, e.g. Hamming, Hanning, Bartlett,...

- It is clear that the resulting spectrum can be written as a frequency domain convolution:

$$\hat{S}_X^{BT}(e^{j\omega T}) = \frac{1}{2\pi} W(e^{j\omega T}) \ast \hat{S}_X(e^{j\omega T})$$

where $W(.)$ is the DTFT of the $2L + 1$ window function and $\hat{S}_X(.)$ is the Periodogram.

- The B-T method can reduce the variance of the periodogram estimate at the expense of some frequency resolution. A special case is the correlogram considered earlier.
• To see this take the expectation

\[ E \left\{ \hat{S}_X^{BT}(e^{j\omega T}) \right\} = \frac{1}{2\pi} W(e^{j\omega T}) \ast E \left\{ \hat{S}_X(e^{j\omega T}) \right\} \]

and there is already “smoothing” in \( E \left\{ \hat{S}_X(e^{j\omega T}) \right\} \)

• To appreciate why windowing controls the variance, we revisit the white Gaussian noise case where the variance of the periodogram could be calculated exactly

• Let \( \{x_n\} \) be a sequence of iid Gaussian random variables with zero mean and variance \( \sigma^2 \) as before

• We can write the BT estimate as

\[
\hat{S}_X^{BT}(e^{j\omega T}) = w_0 \hat{R}_{XX}[0] + \sum_{l=1}^{L} 2w_l \hat{R}_{XX}[l] \cos(\omega Tl)
\]

Note that \( E\{ \hat{S}_X^{BT}(e^{j\omega T}) \} = w_0 E\{ \hat{R}_{XX}[0] \} = w_0 \sigma^2 \)
Squaring gives:

\[ \hat{S}_X^{BT}(e^{j\omega T})^2 = \]

\[ w_0^2 \hat{R}_{XX}[0]^2 + \sum_{l=1}^{L} 4w_l^2 \hat{R}_{XX}[l]^2 \cos(\omega Tl)^2 + \text{cross terms} \]

For this signal

\[ E\left\{ \hat{R}_{XX}[0] \hat{R}_{XX}[l] \right\} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-l-1} E\{x_i^2 x_j x_{j+l}\} \]

\[ = 0, \]

\[ E\left\{ \hat{R}_{XX}[0]^2 \right\} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E\{x_i^2 x_j^2\} \]

\[ = 3\frac{\sigma^4}{N^2}N + \frac{\sigma^4}{N^2}N(N-1) \]

\[ E\left\{ \hat{R}_{XX}[l]^2 \right\} = \frac{1}{N^2} \sum_{i=0}^{N-1-l} E\{x_i^2 x_{i+l}^2\} \]

\[ = \frac{\sigma^4}{N^2}(N - l) \]
The expected value of $\hat{S}_X^{BT}(e^{j\omega T})^2$ is

$$w_0^2\sigma^4\left(\frac{2}{N} + 1\right) + \frac{\sigma^4}{N}\sum_{l=1}^{L} 4w_l^2 \cos(\omega T l)^2 \frac{N - l}{N}$$

as cross terms have zero mean.

Let $L = N - 1$ and subtracting $E\{\hat{S}_X^{BT}(e^{j\omega T})\}^2 = w_0^2\sigma^4$ we see the variance goes to zero as $N \to \infty$ provided $\sum_{l=1}^{\infty} w_l^2 < \infty$.

It is apparent that the window reduces the contribution of $E\{\hat{R}_{XX}[l]\}^2 = \text{var}\{\hat{R}_{XX}[l]\}$ for large $l$ to the variance of $\hat{S}_X^{BT}(e^{j\omega T})$.

For the more general case (Hayes pg. 423)

$$\text{var}\{\hat{S}_X^{BT}(e^{j\omega T})\} \approx S_X(e^{j\omega T})^2 \frac{1}{N} \sum_{l=-L}^{L} w_l^2$$
5 A note on Computation

- The FFT is the natural way to compute all of the spectra required for periodogram, Bartlett method, etc...

- For the data set $\{x_n\}_{n=0}^{N-1}$, the FFT evaluates the spectrum on a discrete grid of frequencies $\omega_p = p2\pi/NT$, $p = 0, 1, \ldots, N - 1$.

- Let $X(e^{j\omega T}) = \sum_{n=0}^{N-1} x_n e^{-j\omega nT}$. Calling the function

$$\text{FFT}\left(\{x_n\}_{n=0}^{N-1}\right)$$

in Matlab will evaluate $X(e^{j\omega T})$ at

$\omega = 0, \frac{2\pi}{NT}, \frac{2\pi}{NT} \times 2, \ldots, \frac{2\pi}{NT}(N - 1)$

- Suppose we require a finer grid of frequencies. Can we still use the FFT function with the same data set?
• Yes - zero-pad \( x_n \) to make up a length \( 2N \) sequence, \( \{x_0, x_1, \ldots, x_{N-1}, 0, \ldots, 0\} \) and call FFT for this sequence.

• This gives an exact evaluation of the DTFT on a grid which is twice as fine:

\[
X_p^{\text{padded}} = \sum_{n=0}^{N-1} x_n e^{-j2np\pi/(2N)}, \quad p = 0, \ldots, 2N-1
\]

where the summation is only over \( 0, \ldots, N-1 \) since the rest of the terms are zero (zero-padding).

• N.B. Zero-padding doesn’t give us better resolution. Think of it more as smoothly ‘joining the dots’ on a coarse grid DFT.