Improving the Periodogram: Data Windows, Lag Windows, and Block Averaging

OR

Classical / Nonparametric Spectrum Estimation

Data Windows

Given an observed signal of length N, a <u>data window</u> is essentially any function that is positive on $0, 1, 2, \ldots, N-1$ and zero elsewhere. Data windows are usually symmetric.

Examples

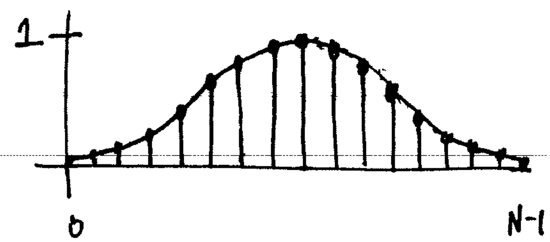
1. Rectangular window:

$$w_R[k] = \begin{cases} 1 & \text{if } 0 \le k \le N-1 \\ 0 & \text{otherwise} \end{cases}$$

2. Bartlett

$$w_B[k] = \begin{cases} \dfrac{N - |k|}{N} & \text{if } 0 \le k \le N-1 \\ 0 & \text{otherwise} \end{cases}$$

3. Hann, Hamming, Parzen, ...



The basic idea is to scale the signal by the window before computing the periodogram:

$\boxed{\text{assume zero mean}}$

$$\hat{P}_{DW}(f) := \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n]\, w[n]\, e^{-2\pi i f k} \right|^2$$

This is called the <u>data-windowed periodogram</u>.

Note that the basic periodogram corresponds to the rectangular window.

Why window? To affect a tradeoff between bias and variance.

Let's see how windowing affects the <u>bias</u> of the periodogram.

Since $w[n] = 0$ for $n < 0$, $n > N-1$, we have

$$\hat{P}_{DW}(f) = \frac{1}{N} \left| \sum_{n=-\infty}^{\infty} x[n]\, w[n]\, e^{-2\pi i f n} \right|^2$$

$$= \frac{1}{N} \left[\sum_{m=-\infty}^{\infty} x[m]\, w[m]\, e^{-2\pi i f m} \right] \left[\sum_{n=-\infty}^{\infty} x[n]\, w[n]\, e^{-2\pi i f n} \right]^*$$

$$= \frac{1}{N} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x[m]\, x^*[n]\, w[m]\, w^*[n]\, e^{-2\pi i f (m-n)}$$

Let $m = n+k$

$$= \frac{1}{N} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x^*[n]\, x[n+k]\, w^*[n]\, w[n+k]\, e^{-2\pi i f k}$$

So the bias is

$$E\{\hat{P}_{DW}(f)\} = \frac{1}{N} \sum_{k} \sum_{n} \gamma_x[k]\, w^*[n]\, w[n+k]\, e^{-2\pi i f k}$$

$$= \sum_{k} \gamma[k] \cdot \underbrace{\left(\frac{1}{N} \sum_{n} w^*[n]\, w[n+k] \right)}_{\text{call this } \phi_w[k]} e^{-2\pi i f k}$$

$$= \sum_{k=-\infty}^{\infty} \gamma_x[k]\, \phi_w[k]\, e^{-2\pi i f k}$$

$$= P_x(f) \star \Phi_w(f)$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_x(u)\, \Phi_w(f-u)\, du$$

Now

$$I_w(f) = \frac{1}{N} \sum_{n=-\infty}^{\infty} w^*[n] \underbrace{\sum_{k=-\infty}^{\infty} w[n+k] e^{-2\pi i f k}}_{W(f) e^{2\pi i f n}}$$

$$= \frac{1}{N} W(f) \cdot \sum_{n=-\infty}^{\infty} w^*[n]\, e^{2\pi i f n}$$

$$= \frac{1}{N} W(f) \cdot W^*(f)$$

$$= \frac{1}{N} |W(f)|^2$$

Conclusion

The bias of data-windowed periodogram is

$$E\{\hat{P}_{DW}(f)\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_x(u) \frac{1}{N}|W(f-u)|^2 du$$

$$= P_x(f) \star \frac{1}{N}|W(f)|^2.$$

Let's check for $w = w_R$:

$$\frac{1}{N}|W_R(f)|^2 = \frac{1}{N}\left(\frac{\sin \pi f N}{\sin \pi f}\right)^2 = W_B(f) \quad \checkmark$$

So in general $\hat{P}_{DW}(f)$ is still biased, and on average it still looks like a "smeared" version of the true PSD.

Is the bias greater or less than it is without windowing?

Kay, p. 70

rectangular window

Hamming window

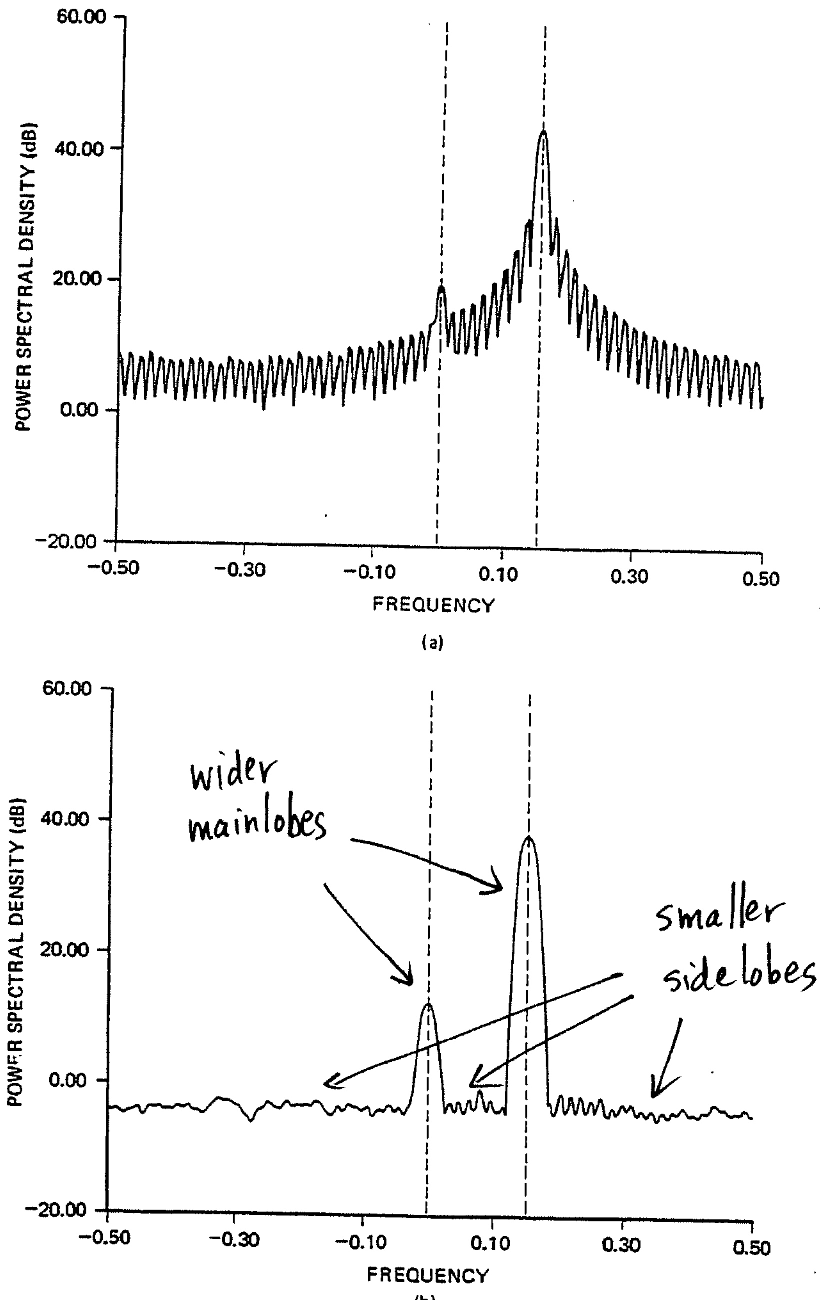


Figure 4.4 Use of data windowing to reduce sidelobes in periodogram of narrowband processes. (a) No data window. (b) Hamming data window.

In general, any window will be more concentrated in time that the rectangular window. By the uncertainty principle, this means that the DTFT of any window will be wider than W_R, and hence the data-windowed periodogram will have an increased bias. This corresponds to a wider mainlobe of the estimated PSD.

In contrast, $\hat{P}_{DW}(f)$ has less variance which can be seen by considering the associated ACV estimate. For example,

$$\hat{\gamma}[N-1] = \underbrace{x[0]\,w[0]}_{\text{usually } 1} \cdot \underbrace{x[N-1]\cdot w[N-1]}_{\text{usually } \leq 1/N}$$

This decreased variance is manifested in smaller sidelobes (less leakage of the main lobe)

Lag Windowing

Recall the periodogram:

$$\hat{P}_{PER}(f) = \sum_{k=-(N-1)}^{N-1} \hat{\gamma}[k] e^{-2\pi i f k}$$

where $\hat{\gamma}$ is the sample autocovariance

$$\hat{\gamma}[k] = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x[n]\, x[n+|k|].$$

$\boxed{\text{for real data}}$

The poor performance of $\hat{P}_{PER}$ can be attributed to the presence of very few lag products for large lags. The idea behind the <u>Blackman-Tukey</u> spectral estimator is to de-emphasize $\hat{\gamma}$ at large lags by using lag windowing:

$$\hat{P}_{BT}(f) = \sum_{k=-(N-1)}^{N-1} w[k]\, \hat{\gamma}[k] e^{-2\pi i f k}$$

Here $w[k]$ is a lag window which satisfies

1) $0 \le w[k] \le w[0] = 1$

2) $w[-k] = w[k]$

3) $w[k] = 0$ for $|k| > M$

where $M \le N-1$. Do not confuse lag windows with data windows, which look the same but are nonzero on $0, \ldots, N-1$.

Special case: $w[k] = w_R[k] = \begin{cases} 1 & \text{if } |k| \le N-1 \\ 0 & \text{if } |k| > N-1 \end{cases}$

$$\Longrightarrow \quad \hat{P}_{BT}(f) = \hat{P}_{PER}(f).$$

TABLE 4.1 COMMON LAG (DATA[a]) WINDOWS

Name	Definition	Fourier transform
Rectangular	$w[k] = \begin{cases} 1 & \lvert k \rvert \le M \\ 0 & \lvert k \rvert > M \end{cases}$	$W(f) = W_R(f) = \dfrac{\sin \pi f(2M+1)}{\sin \pi f}$
Bartlett	$w[k] = \begin{cases} 1 - \dfrac{\lvert k \rvert}{M} & \lvert k \rvert \le M \\ 0 & \lvert k \rvert > M \end{cases}$	$W(f) = W_B(f) = \dfrac{1}{M}\left(\dfrac{\sin \pi f M}{\sin \pi f}\right)^2$
Hanning	$w[k] = \begin{cases} \frac{1}{2} + \frac{1}{2}\cos \dfrac{\pi k}{M} & \lvert k \rvert \le M \\ 0 & \lvert k \rvert > M \end{cases}$	$W(f) = \frac{1}{4} W_R\left(f - \dfrac{1}{2M}\right) + \frac{1}{2} W_R(f) + \frac{1}{4} W_R\left(f + \dfrac{1}{2M}\right)$
Hamming	$w[k] = \begin{cases} 0.54 + 0.46 \cos \dfrac{\pi k}{M} & \lvert k \rvert \le M \\ 0 & \lvert k \rvert > M \end{cases}$	$W(f) = 0.23 W_R\left(f - \dfrac{1}{2M}\right) + 0.54 W_R(f) + 0.23 W_R\left(f + \dfrac{1}{2M}\right)$
Parzen (M even)	$w[k] = \begin{cases} 2\left(1 - \dfrac{\lvert k \rvert}{M}\right)^3 - \left(1 - 2\dfrac{\lvert k \rvert}{M}\right)^3 & \lvert k \rvert \le \dfrac{M}{2} \\ 2\left(1 - \dfrac{\lvert k \rvert}{M}\right)^3 & \dfrac{M}{2} < \lvert k \rvert \le M \\ 0 & \lvert k \rvert > M \end{cases}$	$W(f) = \dfrac{8}{M^3}\left(\dfrac{3}{2}\dfrac{\sin^4 \pi f M/2}{\sin^4 \pi f} - \dfrac{\sin^4 \pi f M/2}{\sin^2 \pi f}\right)$

[a] For the corresponding data window, shift the lag window sequence to the right by M so that $w[k]$ is nonzero over $[0, 2M]$ and replace M by $(N - 1)/2$ (N assumed to be odd). The Fourier transform will be that listed after multiplication by $\exp(-j2\pi f M)$.

Nonnegativity

Unfortunately, unlike $\hat{P}_{PER}$ and $\hat{P}_{DW}$, $\hat{P}_{BT}$ can be negative at some frequencies. Observe

$$\hat{P}_{BT}(f) = \mathrm{DTFT}\left\{ w[k] \cdot \hat{\gamma}[k] \right\}$$

$$= W(f) * \hat{P}_{PER}(f).$$

We know $\hat{P}_{PER}(f) \geq 0$, but for some lag windows, $W(f)$ can be negative for some f.

In general, $W(f) \geq 0$ for all f when the lag window is ________________.

In other words, $w[k]$ needs to be a valid ACV function of some random process.

Among the common lag windows, the Bartlett and Parzen windows satisfy this condition and hence lead to nonnegative spectral estimates.

Bias of Blackman-Tukey

$$E\left\{\hat{P}_{BT}(f)\right\} = E\left\{\sum_{k=-(N-1)}^{N-1} w[k]\,\hat{\gamma}[k]\,e^{-2\pi i f k}\right\}$$

$$= \sum_{-(N-1)}^{N-1} w[k]\; E\left\{\hat{\gamma}[k]\right\} e^{-2\pi i f k}$$

$$= \sum w[k]\cdot\left(w_B[k]\,\gamma[k]\right) e^{-2\pi i f k}$$

w_B: 2N+1 point Bartlett window

$$= W(f) \star W_B(f) \star P_x(f)$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_x(u)\cdot\left[\int_{-\frac{1}{2}}^{\frac{1}{2}} W(v)\,W_B(f-u-v)\,dv\right] du$$

introduces sidelobe vs. mainlobe trade off.

If $N \to \infty$ $\left(W_B(f) \approx \delta(f)\right)$ then

$$E\left\{\hat{P}_{BT}(f)\right\} \approx W(f) \star P_x(f).$$

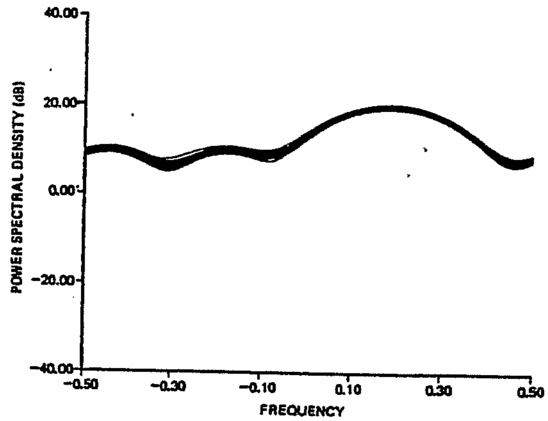
Large M: less bias (narrower main lobe)

Small M: less variance (smaller sidelobes)

Rule of thumb: $M = N/5$

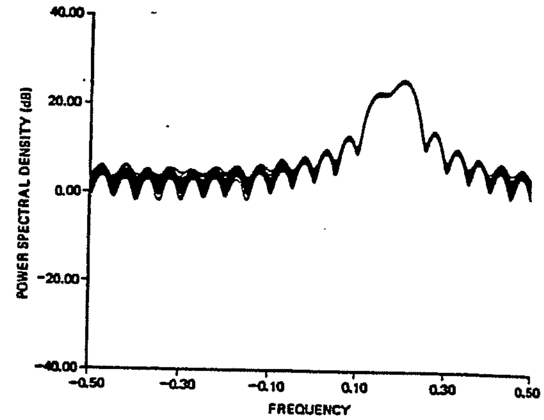
Blackman-Tukey applied to two sinusoids,

lag window $w[k] = w_B[k]$, $M = N/5$



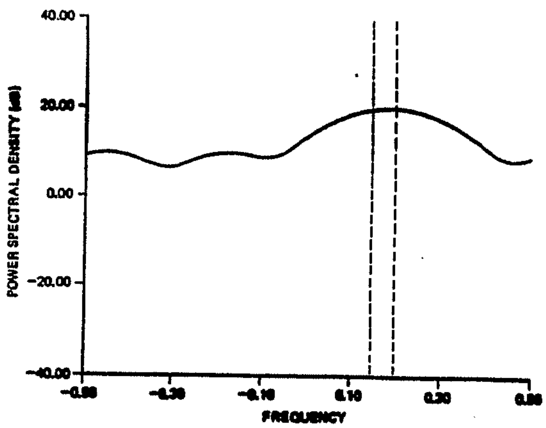
(i) overlaid realizations

(b)

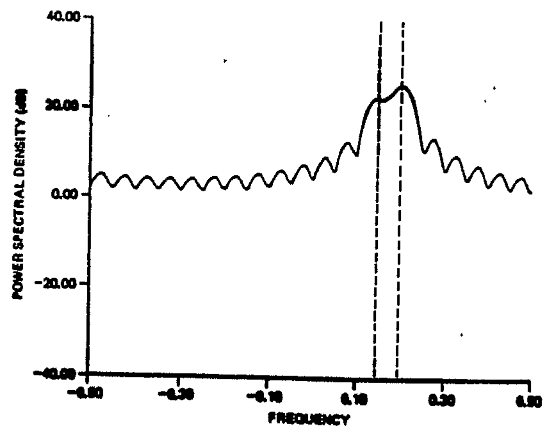


(i) overlaid realizations

(b)



(ii) average of realizations



(ii) average of realizations

$N = 20$

$N = 100$

:) Less variance

:(More bias / less resolution

Block Averaged Periodogram

The formula

$$P_x(f) = \lim_{M\to\infty} E\left\{ \frac{1}{2M+1} \left| \sum_{k=-M}^{M} x[k] e^{-2\pi i f k} \right|^2 \right\}$$

motivated the periodogram

$$\hat{P}_{PER}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-2\pi i f n} \right|^2.$$

The poor performance of $\hat{P}_{PER}$ may also be attributed to a lack of averaging over different realizations (no expectation operator).

Consider an opportunistic setting where we have access to K independent realizations

$$\begin{matrix} x_0[0], & x_0[1], & \ldots, & x_0[L-1] \\ x_1[0], & x_1[1], & \ldots, & x_1[L-1] \\ & \vdots & & \\ x_{K-1}[0], & x_{K-1}[1], & \ldots, & x_{K-1}[L-1] \end{matrix}$$

The <u>averaged periodogram estimator</u> is

$$\hat{P}_{AP}(f) = \frac{1}{K} \sum_{m=0}^{K-1} \hat{P}_{PER}^{(m)}(f)$$

where

$$\hat{P}_{PER}^{(m)}(f) = \frac{1}{L} \left| \sum_{n=0}^{L-1} x_m[n] e^{-2\pi i f n} \right|^2$$

is the basic periodogram on the m^{th} signal.

Recall the following result:

If $Y_0, \ldots, Y_{K-1}$ are independent RVs and $a_0, \ldots, a_{K-1}$ are scalars then

$$E\left\{ \sum_{m=0}^{K-1} a_m Y_m \right\} = \sum_{m=0}^{K-1} a_m E\{Y_m\}$$

and

$$Var\left\{ \sum_{m=0}^{K-1} a_m Y_m \right\} = \sum_{m=0}^{K-1} a_m^2 \, Var\{Y_m\}.$$

Now apply this result with $Y_m = \hat{P}^{(m)}_{PER}(f)$ and $a_m = \frac{1}{K}$ to obtain

$$E\left\{\hat{P}_{AP}(f)\right\} = E\left\{\hat{P}^{(m)}_{PER}(f)\right\} \quad \text{for any } m$$

and

$$\mathrm{Var}\left\{\hat{P}_{AP}(f)\right\} = \frac{1}{K}\mathrm{Var}\left\{\hat{P}^{(m)}_{PER}(f)\right\} \quad \text{for any } m.$$

Conclusions: Averaging

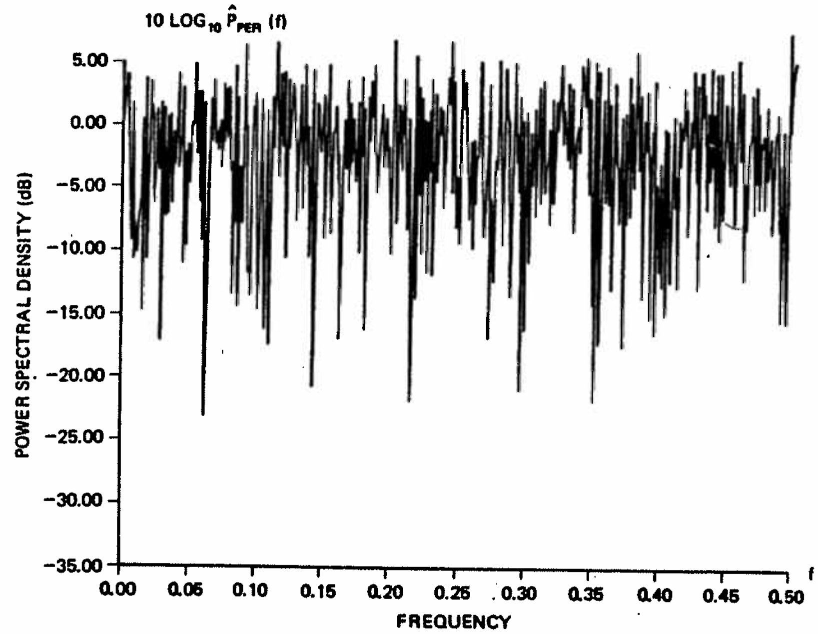
1. Does not change the bias
2. Decreases the variance by a factor of K.

Averaging in Practice

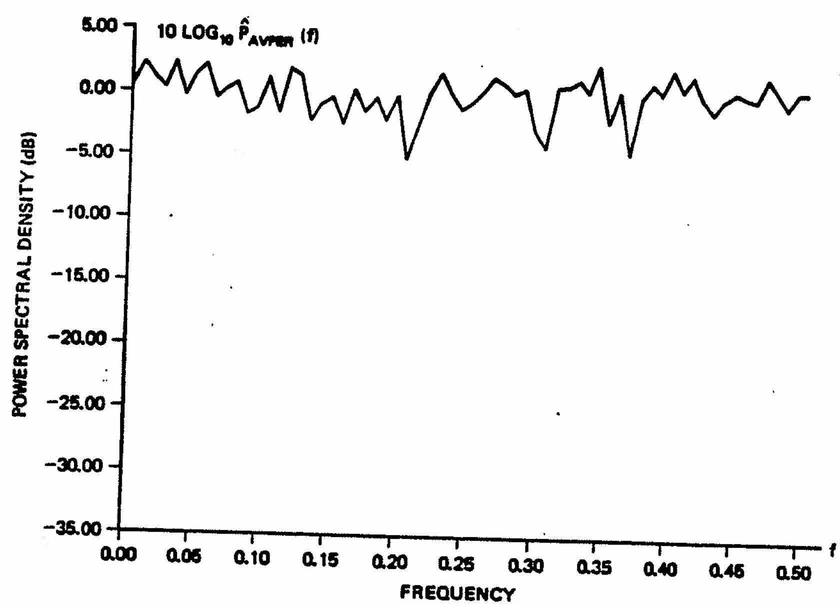
Rarely do we have independent realizations. Rather, we have one realization of length N. In this case we may partition the signal into K nonoverlapping and contiguous blocks of length L, $K \cdot L = N$, and compute $\hat{P}_{AP}$ based on $x_m[n] = x[n + mL]$, $0 \le n \le L-1$, $0 \le m \le K-1$

WGN, $N = 1024$

$L = 1024$
$K = 1$
(basic periodogram)



$L = 128$
$K = 8$

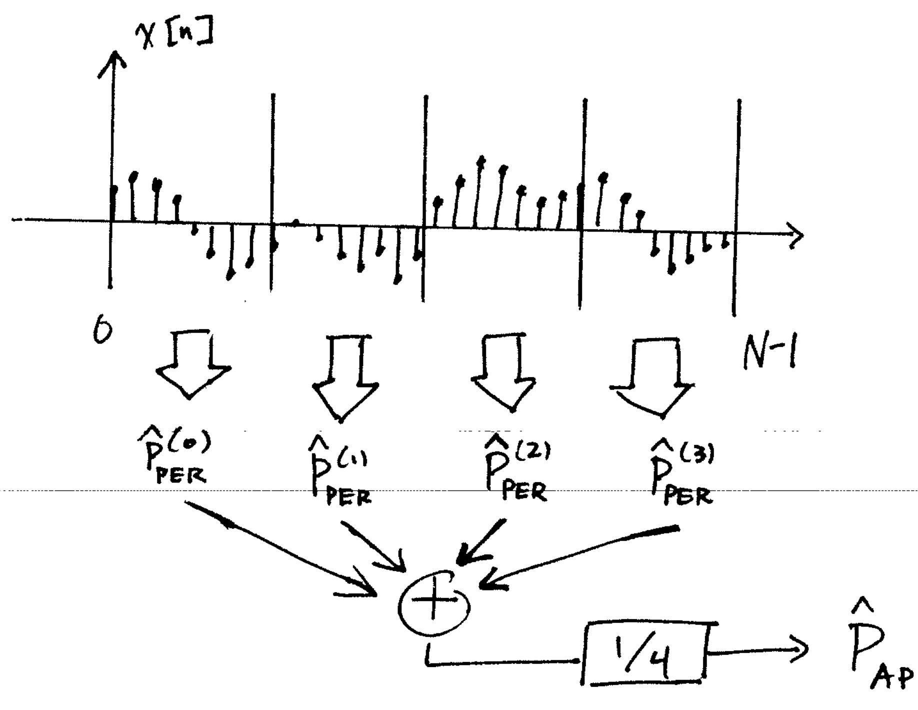


For WGN, the independence assumption actually holds, and we see a major reduction in variance.

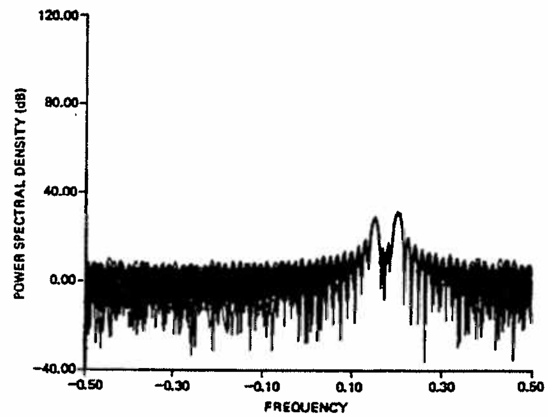
Since the blocks are contiguous, unless the random process is white noise, the blocks are not independent and so the actual variance reduction is less than K. Yet if the ACV decays somewhat rapidly, the independence assumption will hold "approximately" and there will be some variance reduction.

On the other hand, the bias will increase because the block length is $L < N$. Recall the mainlobe of the Fourier transform of a window is wider for a shorter window.

So once again there is a <u>bias-variance tradeoff.</u>

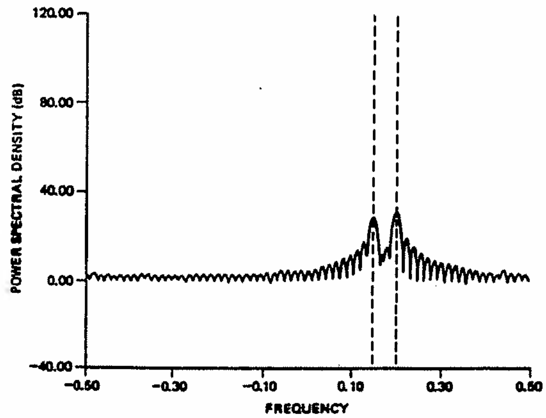


Two sinusoids again, $N=256$

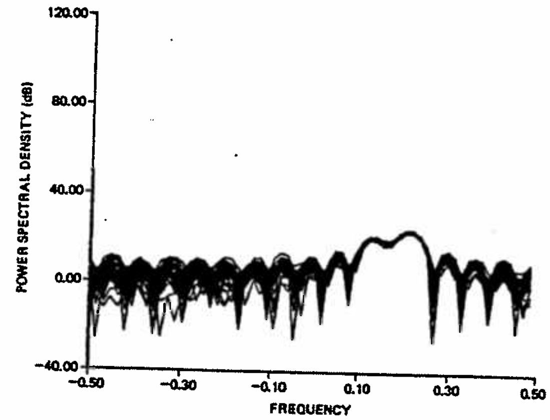


(i) overlaid realizations

(a)

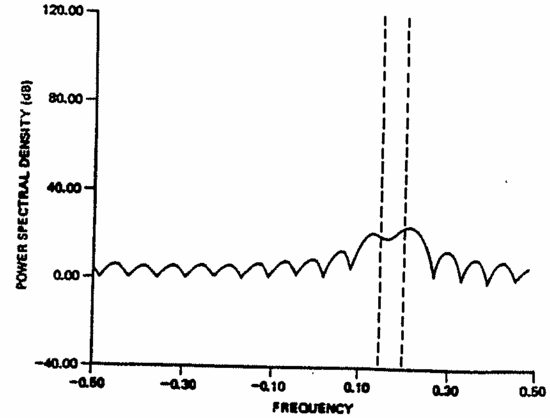


(ii) average of realizations



(i) overlaid realizations

(b)



(ii) average of realizations

$K=4$

$L=64$

$\Rightarrow$ less bias

$K=16$

$L=16$

$\Rightarrow$ less variance

The averaged periodogram is sometimes called the Bartlett periodogram.

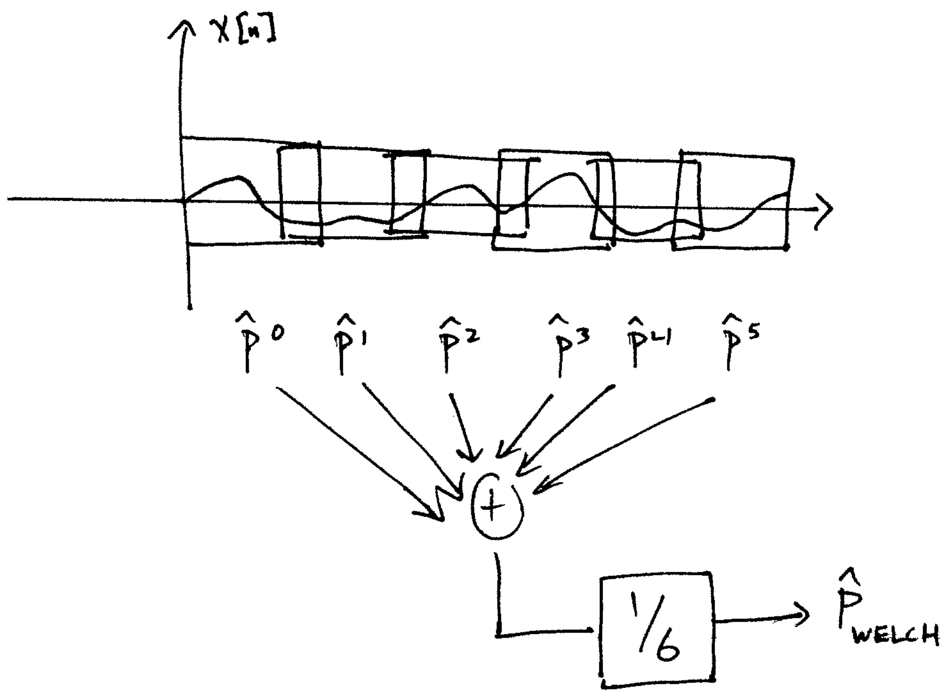
Welch Periodogram

⇨ break data into overlapping blocks

⇨ apply data window to each block

Most popular periodogram method.

Example: 50% overlap with Hann window



The Ultra-Periodogram

Putting it all together: