Practical Classical Methods

Main Problems of the Periodogram

1. Biased Estimate
2. Variance does NOT decrease with increasing N
3. Rapid Fluctuations

All of these arise due to the fact that the periodogram ignores:

- The Expected Value – (It includes no averaging)
- The Limit Operation – (It applies a rectangular window)

in the PSD definition.

Several "classical" methods for partially fixing these have been proposed.

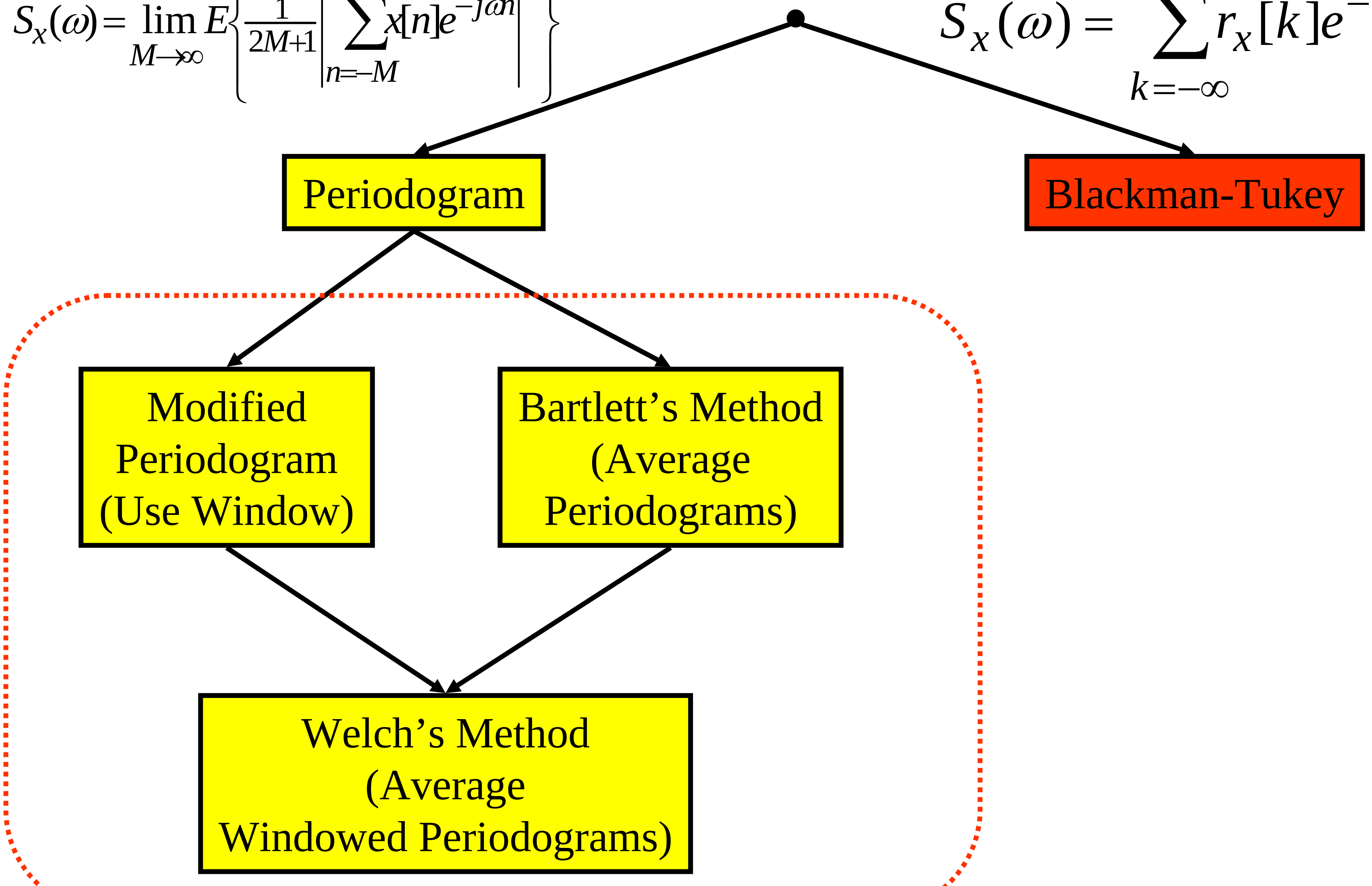
Modifications Based On Periodogram View

- Modified Periodogram
- Bartlett Method
- Welch Method

Recall: Family of "Classical" Methods

$$S_x(\omega) = \lim_{M\to\infty} E\left\{ \frac{1}{2M+1} \left| \sum_{n=-M}^{M} x[n] e^{-j\omega n} \right|^2 \right\}$$

$$S_x(\omega) = \sum_{k=-\infty}^{\infty} r_x[k] e^{-j\omega k}$$



Modified Periodogram - Windowed

H-8.2.3

The "Modified Periodogram" uses a non-rectangular window.

Motivated by:

$$\boxed{E\left\{\hat{S}_{PER}(\omega)\right\} = \frac{1}{2\pi} S_x(\omega) * W_B(\omega)}$$

We see that it is this convolution that keeps the periodogram from being unbiased. And… we recognize that to get unbiased performance we would need $W_B(\omega) = \delta(\omega)$.

As our previous studies of windows have shown, that is impossible. But we can **choose non-rectangular windows** to reduce the sidelobe leakage.

This reduces the bias effect.

But… at the expense of degraded resolution.

Example: Two Sinusoids – **See Hayes Fig. 8.10**

Modified Periodogram - Definition

The “Modified Periodogram” uses a non-rectangular window and therefore has to be scaled to account for the loss of power due to the window. This scaling is required to make the Modified Periodogram asymptotically unbiased:

$$\boxed{\hat{S}_{MP}(\omega) = \frac{1}{NU}\left|\sum_{n=0}^{N-1} x[n]w[n]e^{-j\omega n}\right|^2}$$

where the scaling factor is $U = \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1}\left|w[n]\right|^2$

<<Note: U = 1 for a rectangular window>>

As in the ordinary periodogram, the DFT/FFT is used for computation and zero-padding is usually used.

Modified Periodogram - Performance

The Modified Periodogram:

- Has reduced bias but is still biased.
- Is asymptotically unbiased.
- Has variance that roughly equals that of the periodogram.

Bartlett's Method: Averaged Periodogram

H-8.2.4

One of the main flaws in the periodogram is the lack of averaging.

<< See how ensemble averaging improves it… Hayes Fig. 8.8>>

This lack of averaging is what leads to the non-decreasing variance as well as the rapid fluctuations of the periodogram.

Now… in practice we have only one realization…

So what do we do to allow averaging????

We HOPE that the process is ergodic!!!!

A process is **ergodic** if time averaging of any realization is equivalent to ensemble averaging.

Bartlett's Method – Definition

The signal data of length N is chopped into K non-overlapping blocks of length L (the length L is a "design choice"); $N = KL$:

$$\text{Block Definition:} \quad x_i[n] = x[n + iL] \quad \begin{array}{l} n = 0,1, \cdots, L-1 \\ i = 0,1, \cdots, K-1 \end{array}$$

$$\begin{aligned} \hat{S}_B(\omega) &= \frac{1}{K} \sum_{i=0}^{K-1} \frac{1}{L} \left| \sum_{n=0}^{L-1} x_i[n] e^{-j\omega n} \right|^2 \\ &= \frac{1}{N} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} x_i[n] e^{-j\omega n} \right|^2 \end{aligned}$$

Bartlett's Method – Variance Improvement

The intent of averaging here is to improve the variance. To see how lets first just look at a simple related example:

Let X_i, $i = 0, 1, \ldots, K$-1 be a sequence of independent, identically distributed RVs each having zero-mean and variance σ^2. What is the variance of the "data analysis average" of them?

$$\boxed{\overline{X} = \frac{1}{K}\sum_{i=0}^{K-1} X_i}$$

$$\begin{aligned}
\operatorname{var}\{\overline{X}\} &= E\{\overline{X}^2\} - \underbrace{\left[E\{\overline{X}\}\right]^2}_{=0} \\
&= \frac{1}{K^2}\sum_{i=0}^{K-1}\sum_{k=0}^{K-1} E\{X_i X_k\} \\
&= \frac{1}{K^2}\sum_{i=0}^{K-1}\sum_{k=0}^{K-1} \sigma^2 \delta[i-k] \\
&= \frac{K\sigma^2}{K^2} = \frac{\sigma^2}{K}
\end{aligned}$$

"Data Averaging" *K* RV's together reduces the variance by a factor of *K*!!!!

<u>Bartlett's Method – Variance Imp. (cont.)</u>

Since Bartlett's Method "data averages" K periodograms together we should be able to use this result… <u>IF</u> the individual periodograms are independent (stronger than uncorrelated). But they aren't!!!! But since the blocks do not overlap there is likely to be only a small amount of correlation…. so under this simplification (and for the white-noise assumption made earlier):

$$\boxed{\operatorname{var}\left\{\hat{S}_B(\omega)\right\} \approx \frac{1}{K} S_x^2(\omega)}$$

Thus, the more blocks used, the better the variance of the estimate!!

But, for a given data length N,

More Blocks means Shorter Blocks

➔ Shorter Blocks means Poorer Resolution

Fundamental Trade Between Variance and Resolution

Bartlett's Method – Examples

1. White Noise: **See** **Hayes Fig. 8.14**
   Notice the reduced fluctuation with increasing *K*

2. Two Sinusoids in White Noise: **See** **Hayes Fig. 8.15**
   Notice that as K increases:
   - The Fluctuations Decrease
   - The Resolution Gets Worse

Welch's Method: Averaged Windowed P'Gram

H-8.2.5

We've seen:

Windowing helps the Bias

Averaging helps the Variance

Welch: Do Both!!!! And…. use Overlapped Blocks

Block Definition: $x_i[n] = x[n + iD] \quad n = 0,1, \cdots, L-1$

$i = 0,1, \cdots, K-1$

The amount of overlap is $L - D$ points:

$D = L$:	No Overlap
$D = L/2$:	50% Overlap (Most Common)
$D = 3L/4$:	25% Overlap

$$\boxed{\hat{S}_W(\omega) = \frac{1}{KLU} \sum_{i=0}^{K-1} \left| \sum_{n=0}^{L-1} x_i[n] w[n] e^{-j\omega n} \right|^2}$$

Implement Using DFT/FFT & Zero-Padding

Welch's Method: Variance

Analysis is beyond scope of this class.

Variance "has been shown to be" for 50% overlap:

$$\boxed{\operatorname{var}\left\{\hat{S}_W(\omega)\right\} \approx \frac{9}{16}\frac{L}{N} S_x^2(\omega)}$$

Compared to Bartlett's method (No Overlap) for the same *N* and *L:*

$$\boxed{\operatorname{var}\left\{\hat{S}_W(\omega)\right\} \approx \frac{9}{16} \operatorname{var}\left\{\hat{S}_B(\omega)\right\}}$$

almost a 50% reduction!!!

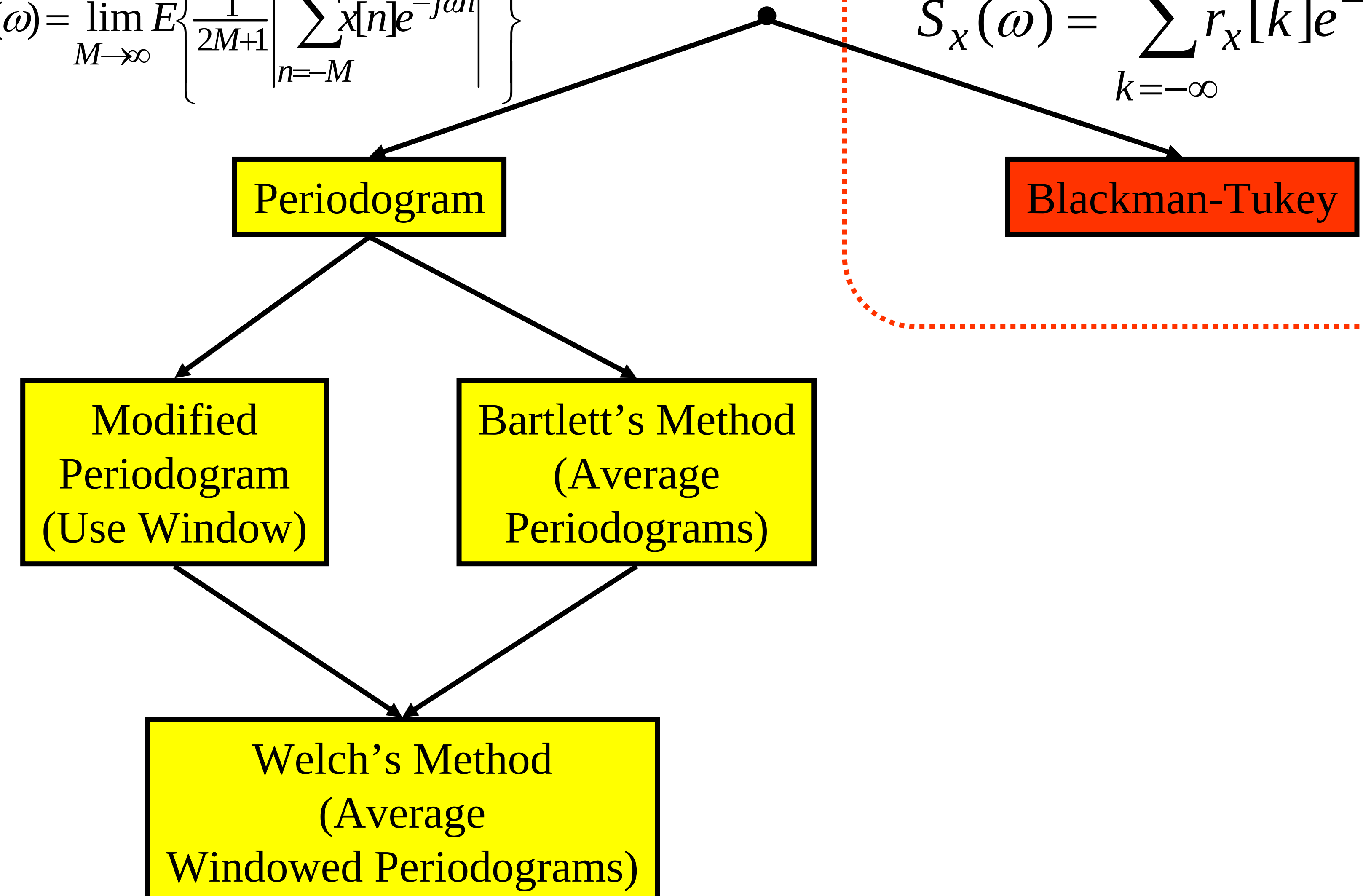
Modifications Based on ACF View

- Blackman-Tukey Method

Recall: Family of "Classical" Methods

$$S_x(\omega) = \lim_{M \to \infty} E\left\{ \frac{1}{2M+1} \left| \sum_{n=-M}^{M} x[n] e^{-j\omega n} \right|^2 \right\}$$

$$S_x(\omega) = \sum_{k=-\infty}^{\infty} r_x[k] e^{-j\omega k}$$



Recall Periodogram-Based Methods

Periodogram's biggest problem is a variance that does not decrease with increasing *N*

The methods we've seen dealt with this by averaging.

There is another way to combat this… to see how, we need to write the periodogram differently – motivated by the Wiener-Khinchine Theorem:

$$S_x(\omega) = \sum_{k=-\infty}^{\infty} r_x[k] e^{-j\omega k}$$

$$\text{ACF: } r_x[k] = E\left\{ x[n] x^*[n+k] \right\}$$

Periodogram Re-Interpreted via WK Theorem

The Periodogram can be written as:

$$\hat{S}_{PER}(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j\omega n}\right|^2 = \frac{1}{N}\left[\sum_{m=0}^{N-1} x[m]e^{-j\omega m}\right]\left[\sum_{n=0}^{N-1} x[n]e^{-j\omega n}\right]^*$$

$$= \frac{1}{N}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x[m]x^*[n]e^{-j\omega(m-n)}$$

Let $\quad m = n + k \quad \Rightarrow \quad k = m - n \quad \Rightarrow \quad k \in [-(N-1),(N-1)]$

$$= \frac{1}{N}\sum_{k=-(N-1)}^{N-1}\ \sum_{n=0}^{N-1} x[n+k]x^*[n]e^{-j\omega k}$$

Note: $x[n+k] = 0$ for $n + k > N-1$ i.e. for $n > N - k - 1$

$$= \sum_{k=-(N-1)}^{N-1} \underbrace{\left[\frac{1}{N}\sum_{n=0}^{N-|k|-1} x[n+k]x^*[n]\right]}_{=\hat{r}_b[k]} e^{-j\omega k}$$

<u>Periodogram Re-Interpreted (cont.)</u>

Thus…The Periodogram can be written as:

$$\boxed{\hat{S}_{PER}(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{r}_b[k] e^{-j\omega k}}$$

i.e…. as a DTFT of an <u>*estimated* ACF</u> given by

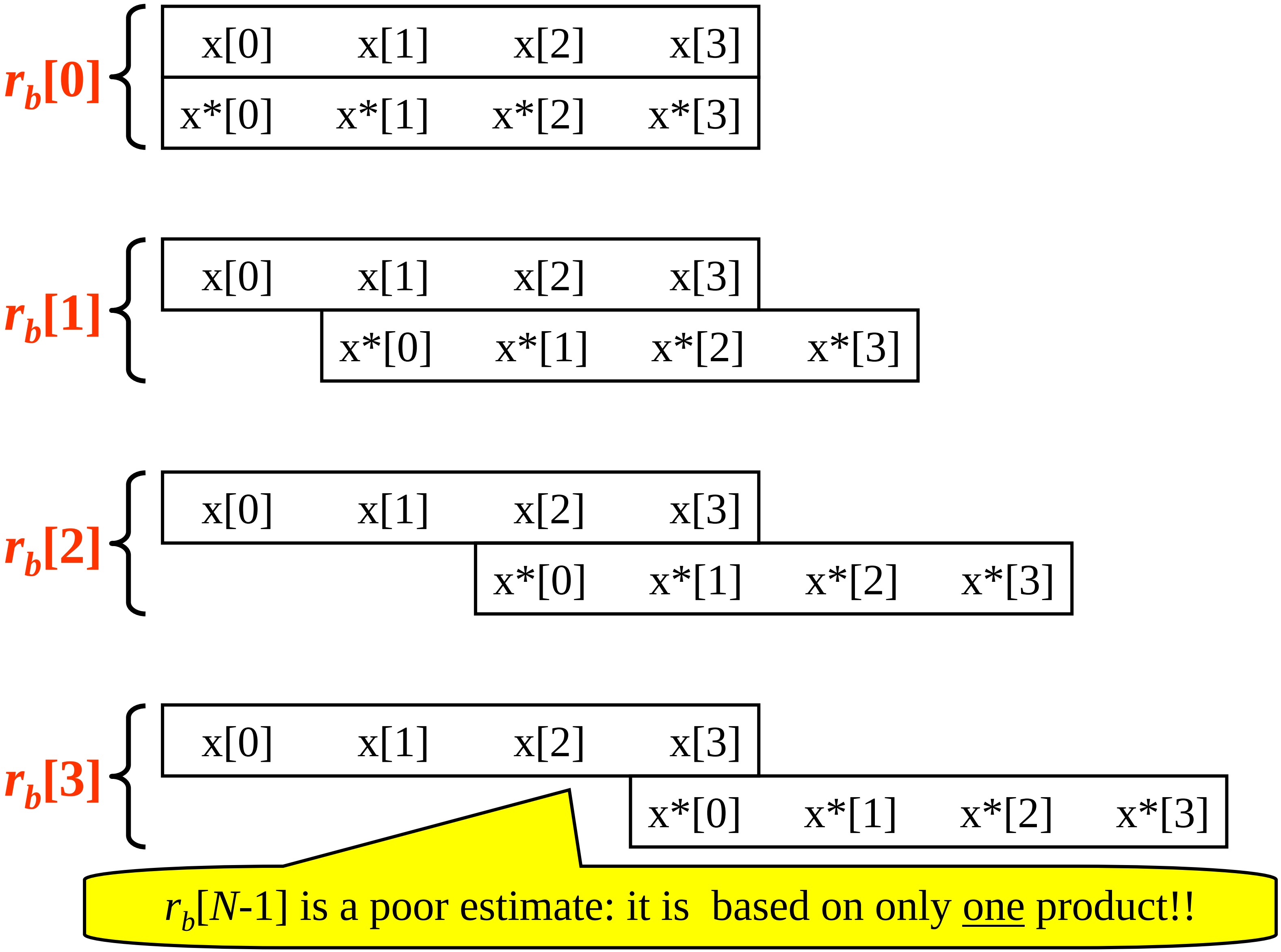
$$\boxed{\hat{r}_b[k] = \tfrac{1}{N} \sum_{n=0}^{N-|k|-1} x[n+k]x^*[n]}$$

The subscript "b" on the ACF estimate indicates that this is a <u>biased estimate</u> of the ACF.

It is the poor quality of this ACF estimate that gives rise to the periodogram's poor quality!!!!

Aha… New Insight!!!!

ACF Estimation



Blackman-Tukey Method - Defined

For the biased ACF estimate, the estimated ACF "lags" for large $|k|$ values are unreliable! **What can we do to fix this???**

De-emphasize these unreliable "lags" by applying a window to the biased ACF estimate. This is the Blackman-Tukey Method:

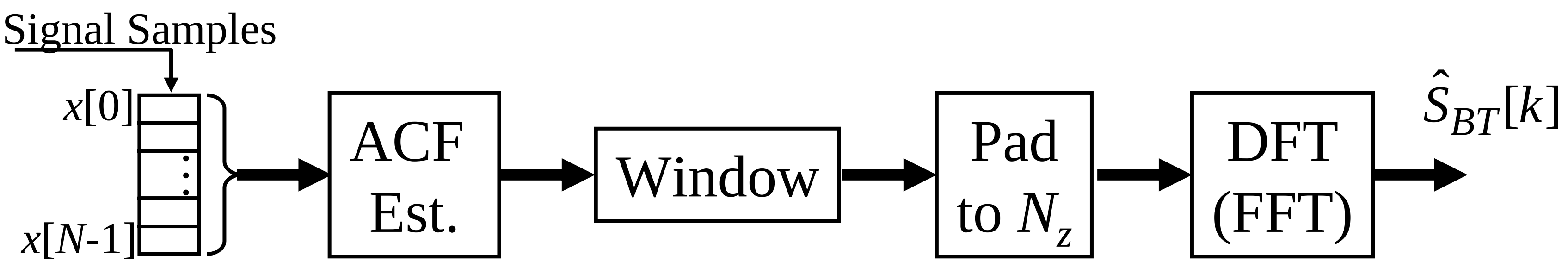
$$\boxed{\hat{S}_{BT}(\omega) = \sum_{k=-M}^{M} w[k]\hat{r}_b[k]e^{-j\omega k} \quad \text{with } M < N-1}$$

Means that we don't even use some of the possible lag estimates

Since windows taper off to zero at their edges this causes the poor-quality estimates at large $|k|$ values to have less impact on the PSD estimate.

Blackman-Tukey - Computation

In practice we compute this using the DFT(FFT) (usually using zero-padding) – which computes the DTFT at discrete frequency points ("DFT Bins"):



Blackman-Tukey – Freq. Domain Interp

Although we always implement the BT Method as just shown, it is useful to explore a frequency domain interpretation of it. By using the multiplication-convolution theorem for DTFT we have:

(Product in Time Domain) ⇔ (Convolution in Frequency Domain)

$$\hat{S}_{BT}(\omega) = \sum_{k=-M}^{M} w[k]\hat{r}_b[k]e^{-j\omega k}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega-\xi)\hat{S}_{PER}(\xi)d\xi = \frac{1}{2\pi} W(\omega) \underset{circ}{*} \hat{S}_{PER}(\omega)$$

The BT Estimate is a <u>smoothed version</u> of the Periodogram.

This is why BT gets rid of the
variance problem of the periodogram!!!

Blackman-Tukey vs. Welch/Bartlett Method

Both the BT method and the Welch/Bartlett method are successful in reducing the variance compared to the pure Periodogram. But **HOW** they do it is quite different!

• Welch/Bartlett does it by averaging away the variations over many computed periodograms.

• Blackman-Tukey does it by smoothing the variations out of a single periodogram.

<u>Blackman-Tukey – Performance - Bias</u>

So far we've alluded to the fact that BT improves upon the periodogram… but of course we need to **<u>PROVE</u>** it!!

<u>Bias</u>

$$E\left\{\hat{S}_{BT}(\omega)\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega-\xi)\ \underbrace{E\left\{\hat{S}_{PER}(\xi)\right\}}_{\approx S_x(\xi)\text{ for large }N}\ d\xi$$

Since Asymp. Unbiased

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega-\xi) S_x(\xi) d\xi$$

$$\boxed{E\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{2\pi} W(\omega) \underset{circ}{*} S_x(\omega)}$$

Blackman-Tukey – Performance - Variance

$$\operatorname{var}\left\{\hat{S}_{BT}(\omega)\right\} = E\left\{\left[\hat{S}_{BT}(\omega) - E\left\{\hat{S}_{BT}(\omega)\right\}\right]^2\right\}$$

$$= \frac{1}{4\pi^2} E\left\{\left[\int_{-\pi}^{\pi} W(\omega-\xi)\hat{S}_{PER}(\xi)d\xi - \int_{-\pi}^{\pi} W(\omega-\xi)E\left\{\hat{S}_{PER}(\xi)\right\}d\xi\right]^2\right\}$$

$$= \frac{1}{4\pi^2} E\left\{\left[\int_{-\pi}^{\pi} W(\omega-\xi)\left[\hat{S}_{PER}(\xi) - E\left\{\hat{S}_{PER}(\xi)\right\}\right]d\xi\right]^2\right\}$$

$$= \frac{1}{4\pi^2} E\left\{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} W(\omega-\xi)W(\omega-\lambda)\left[\hat{S}_{PER}(\xi) - E\left\{\hat{S}_{PER}(\xi)\right\}\right]\left[\hat{S}_{PER}(\lambda) - E\left\{\hat{S}_{PER}(\lambda)\right\}\right]d\xi d\lambda\right\}$$

$$= \frac{1}{4\pi^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} W(\omega-\xi)W(\omega-\lambda)\underbrace{E\left\{\left[\hat{S}_{PER}(\xi) - E\left\{\hat{S}_{PER}(\xi)\right\}\right]\left[\hat{S}_{PER}(\lambda) - E\left\{\hat{S}_{PER}(\lambda)\right\}\right]\right\}}_{=\operatorname{cov}\left\{\hat{S}_{PER}(\xi),\hat{S}_{PER}(\lambda)\right\}}d\xi d\lambda$$

$$= \frac{1}{4\pi^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} W(\omega-\xi)W(\omega-\lambda)\operatorname{cov}\left\{\hat{S}_{PER}(\xi),\hat{S}_{PER}(\lambda)\right\}d\xi d\lambda$$

Have Approx. Result for Non-White Case

Blackman-Tukey – Performance - Variance

$$\operatorname{var}\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} W(\omega-\xi)W(\omega-\lambda)S_x(\xi)S_x(\lambda)\left(\frac{\sin[N(\xi-\lambda)/2]}{N\sin[(\xi-\lambda)/2]}\right)^2 d\xi d\lambda$$

Now, further approximation must be done do get a meaningful result. If *N* is large enough, the "sin-over-sin" kernel will be enough like a delta function (with area $2\pi/N$) to treat it approximately as one:

$$\operatorname{var}\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{2\pi N}\int_{-\pi}^{\pi} W^2(\omega-\lambda)S_x^2(\lambda)d\lambda$$

Now, further approximation: assume that the true PSD is fairly constant over any interval of width = to mainlobe of $W(\omega)$:

$$\operatorname{var}\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{2\pi N}S_x^2(\omega)\int_{-\pi}^{\pi} W^2(\omega-\lambda)d\lambda = S_x^2(\omega)\frac{1}{N}\sum_{k=-M}^{M} w^2[k]$$

Use Parseval's Theorem

Blackman-Tukey – Performance Insight

$$E\left\{\hat{S}_{BT}(\omega)\right\} \approx \frac{1}{2\pi} W(\omega) \underset{circ}{*} S_x(\omega)$$ Bias

$$\operatorname{var}\left\{\hat{S}_{BT}(\omega)\right\} \approx S_x^2(\omega)\left[\frac{1}{N}\sum_{k=-M}^{M} w^2[k]\right]$$ Variance

Basic Tradeoff Between Bias and Variance:

- Need Large M to get small bias
  - In order to get narrow mainlobe and low sidelobes
- Need $M << N$ to get low variance
  - In order to reduce the bracketed term in variance Eq.

Recommended: $M < N/5$

Performance Comparison for Classical Methods

Performance Measures

We've seen that we care about three main things:

1. Bias
2. Variance
3. Resolution

Combined into "Variability" – see below

… and there is usually a tradeoff between them – especially between variance & resolution.

It is desirable to come up with a single-measure way to compare the methods:

Figure-of-Merit = (Variability)×(Resolution)

Performance Measures - Variability

As we've seen, variance is an important quality measure for PSD estimation. However, by itself it tells very little about quality: large variance in an estimate of a large number may be better than medium variance in an estimate of a small number. Thus we need a way to normalize the variance:

This is called **Variability**:

$$\boxed{v = \frac{\mathrm{var}\{\hat{S}(\omega)\}}{E^2\{\hat{S}(\omega)\}}}$$

Note that variability is a unitless quantity.

Small *v* is Desirable

Performance Measures - Resolution

As we saw in our studies of Ch. 6 in Porat, one of the important measures of goodness for spectral analysis is resolution – the ability to see two closely-spaced sinusoids.

Recall: The width of the mainlobe of the window's kernel impacts this ability. There are many ways to measure resolution – Hayes defines resolution as:

$$\boxed{\Delta\omega = 6\,\text{dB Width of Mainlobe}}$$

Small Δω is Desirable

Recall – Two things impact ML Width:

1. Window Length: $\Delta\omega\downarrow$ as Length $\uparrow$
2. Window Shape (e.g. Hanning, Hamming, Etc.)

Recall – There is a tradeoff between $\Delta\omega$ and SL level

Overall Figure of Merit

It is helpful to have a single measure by which to compare methods. This is done using the following Figure of Merit:

$$\boxed{\mathcal{M} = v \times \Delta\omega}$$

Since v and $\Delta\omega$ are both required to be as small as possible, we also want the figure of merit $\mathcal{M}$ to be as small as possible.

Performance - Periodogram

Using our results for bias and variance of the periodogram:

$$\boxed{v_{PER} = \frac{\mathrm{var}\{\hat{S}_{PER}(\omega)\}}{E^2\{\hat{S}_{PER}(\omega)\}} = \frac{S_x^2(\omega)}{S_x^2(\omega)} = 1}$$

Recalling that $E\left\{\hat{S}_{PER}(\omega)\right\} = \frac{1}{2\pi} S_x(\omega) \underset{circ}{*} W_B(\omega)$

we need to assess resolution based on a Bartlett Window:

$$\boxed{\Delta\omega_{PER} = 0.89 \frac{2\pi}{N}}$$

Thus, the periodogram's figure of merit is:

$$\boxed{\mathcal{M}_{PER} = 0.89 \frac{2\pi}{N}}$$

Performance – Bartlett's Method

For N samples, use K blocks of length L where $N = KL$

A reduction in variance is achieved by averaging over K Blocks:

$$\boxed{v_B = \frac{\operatorname{var}\{\hat{S}_B(\omega)\}}{E^2\{\hat{S}_B(\omega)\}} \approx \frac{\frac{1}{K}S_x^2(\omega)}{S_x^2(\omega)} = \frac{1}{K} = \frac{L}{N} < \left(v_{PER} = 1\right)}$$

<< Using More Blocks Improves Variance>>

Since we are using blocks of length $L = N/K$ the resolution is

$$\boxed{\Delta\omega_B = 0.89\frac{2\pi}{L} = 0.89K\frac{2\pi}{N} = K\Delta\omega_{PER}}$$

<< But… Using More Blocks Degrades Resolution>>

Thus, the Bartlett's Method figure of merit is:

$$\boxed{\mathcal{M}_B = 0.89\frac{2\pi}{N} = \mathcal{M}_{PER}}$$

Performance – Welch's Method w/ 50% Overlap

For N samples, use overlapped blocks of length L

A reduction in variance is achieved by averaging over Blocks:

$$\boxed{v_W = \frac{\operatorname{var}\{\hat{S}_W(\omega)\}}{E^2\{\hat{S}_W(\omega)\}} = \frac{9}{16}\frac{L}{N} < \left(v_B = \frac{L}{N}\right) < \left(v_{PER} = 1\right)}$$

<< Overlapping Gives More Blocks & Improves Variance>>

Consider using a Bartlett window (remember – this is applied directly to the data so you actually get "double application").

Since we are using blocks of length L the resolution is

$$\boxed{\Delta\omega_W = 1.28\frac{2\pi}{L} > \Delta\omega_B}$$

Other Windows Give Different Values

Thus, the Welch's Method figure of merit is:

$$\boxed{\mathcal{M}_W = 0.72\frac{2\pi}{N} < \mathcal{M}_{PER} = \mathcal{M}_B}$$

Performance – BT Method

Consider using a Bartlett window on the estimated ACF. The window length is 2M where M << N

$$\boxed{\nu_{BT} = \frac{\text{var}\{\hat{S}_{BT}(\omega)\}}{E^2\{\hat{S}_{BT}(\omega)\}} = \frac{2M}{3N}}$$

<< Using shorter window improves variance>>

The effect is a "double application" of the 2M-length window:

$$\boxed{\Delta\omega_{BT} = 1.28\frac{2\pi}{2M} = 0.64\frac{2\pi}{M}}$$

<< Using shorter window degrades resolution>>

Thus, the BT Method figure of merit is:

$$\boxed{\mathcal{M}_{BT} = 0.43\frac{2\pi}{N} < \mathcal{M}_W < \mathcal{M}_{PER} = \mathcal{M}_B}$$

Performance Comparison of Classical Methods

	Variability ν	Resolution $\Delta\omega$	Merit $\mathcal{M}$
Periodogram	1	$0.89\frac{2\pi}{N}$	$0.89\frac{2\pi}{N}$
Bartlett	$\frac{L}{N}$	$0.89\frac{2\pi}{L}$	$0.89\frac{2\pi}{N}$
Welch (50% Overlap & Bartlett Window)	$\frac{9}{16}\frac{L}{N}$	$1.28\frac{2\pi}{L}$	$0.72\frac{2\pi}{N}$
Blackman-Tukey	$\frac{2}{3}\frac{M}{N}$	$0.64\frac{2\pi}{M}$	$0.43\frac{2\pi}{N}$

=

Other Windows Can be Used

Trade-Off Between Variance & Resolution

Res ~ 1/(DFT Size)

Merit ~ 1/*N*

Complexity Comparison of Classical Methods

Welch and BT methods are the most commonly used ones. But counting the number of complex multiplies needed for each one, it is easy to see that:

Welch requires a bit more computation then BT

BUT… bear in mind:

For BT, none of the ACF lags can be estimated until ALL of the data is obtained – therefore no computing can be done until all the data is obtained

For Welch, DFT’s can be started as soon as each block arrives.

Welch <u>MIGHT</u> have a real-time advantage!!!