Improved periodogram-based methods

Motivation

- Periodogram: based on the FFT $\rightarrow$ simple, computationally efficient PSD estimate.
- Asymptotically unbiased but inconsistent estimator: variance does not improve with observed sample size N.
- Periodogram variance mainly caused by ACS estimates with large $|k|$, which have large variance due to small number of contributing samples.
- To improve variance, two main families of approaches:
  - **periodogram smoothing**: Blackman-Tuckey, Daniell
  - **periodogram averaging**: Bartlett, Welch.

Blackman-Tuckey method: periodogram smoothing

- ACS estimates with large lag $|k|$ have high variance $\rightarrow$ increased variance of PSD estimate.
- **Idea**: apply small weights to covariances $\hat{r}(k)$ with large $|k|$ to reduce variance.
- **Drawback**: smaller number of ACS estimates $\rightarrow$ resolution reduction.

Blackman-Tuckey method

1. Compute an **estimate** $\hat{r}(k)$ of the **ACS**, for $-(M-1) \leq k \leq (M-1)$.
2. Choose a **window sequence** $w(k)$ of length $(2M-1)$ samples.
3. Compute the **Blackman-Tuckey spectral estimate**:

$$\hat{S}_{\mathrm{BT}}(\omega) = \sum_{k=-(M-1)}^{M-1} w(k)\hat{r}(k)\mathrm{e}^{-\jmath\omega k}.$$

- $M \leq N$ — otherwise, no variance reduction over periodogram.
- $\hat{S}_{\mathrm{BT}}(\omega) =$ DTFT of the product of $w(k)$ and $\hat{r}(k)$ — for biased ACS estimate:

$$\hat{S}_{\mathrm{BT}}(\omega) = \hat{S}_{\mathrm{P}}(\omega) * W(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{S}_{\mathrm{P}}(\xi)W(\omega-\xi)d\xi.$$

$$\mathrm{E}\{\hat{S}_{\mathrm{BT}}(\omega)\} = \mathrm{E}\{\hat{S}_{\mathrm{P}}(\omega)\} * W(\omega) = \underbrace{S(\omega) * W_{\mathrm{B}}(\omega)}_{\mathrm{E}\{\hat{S}_{\mathrm{P}}(\omega)\}} * W(\omega) \qquad \text{with } W(\omega) = \mathcal{F}\{w(k)\}.$$

Blackman-Tuckey method — properties

- $\hat{S}_{\mathrm{BT}}(\omega)$ is a **"locally" smoothed periodogram** $\rightarrow$ effects:
  - variance decreases substantially
  - bias increases slightly $\rightarrow$ spectral resolution worsens.
- If $W(\omega) \geq 0 \quad \Rightarrow \quad \hat{S}_{\mathrm{BT}}(\omega) \geq 0$: **BT spectral estimate is nonnegative**.
- **Equivalent time width**

$$N_e = \frac{1}{w(0)} \sum_{k=-(M-1)}^{M-1} w(k) = O(M)$$

- **Equivalent bandwidth**

$$\beta_e = \frac{1}{2\pi W(0)} \int_{-\pi}^{\pi} W(\omega) d\omega$$

- **Time-bandwith product** (shown in tutorial)

$$\boxed{N_e \beta_e = 1}$$

Blackman-Tuckey method — properties (cont'd)

Mean and variance for $N \gg M$:

$$\mathrm{E}\{\hat{S}_{\mathrm{BT}}(\omega)\} = \mathrm{E}\{\hat{S}_{\mathrm{P}}(\omega)\} * W(\omega) = \underbrace{S(\omega) * W_{\mathrm{B}}(\omega)}_{\mathrm{E}\{\hat{S}_{\mathrm{P}}(\omega)\}} * W(\omega) \underset{N \gg M}{\underset{\uparrow}{\approx}} S(\omega) * W(\omega)$$

$$\mathrm{var}\{\hat{S}_{\mathrm{BT}}(\omega)\} \approx S^2(\omega) \frac{1}{N} \sum_{k=-(M-1)}^{M-1} w^2(k)$$

Resolution–variance tradeoff

- spectral resolution is proportional to $\beta_e = 1/N_e = O(1/M)$
- variance is proportional to M/N.

BT method design

1. **Choose window length parameter** M as a tradeoff between variance and bias
   - bias decreases with M
   - variance increases with M.
2. **Choose window shape** to compromise between smearing and leakage
   - main lobe width determines smearing
   - sidelobe level determines leakage
   - energy in main lobe and sidelobes cannot be reduced simultaneously once M is chosen.

Common windows and their properties

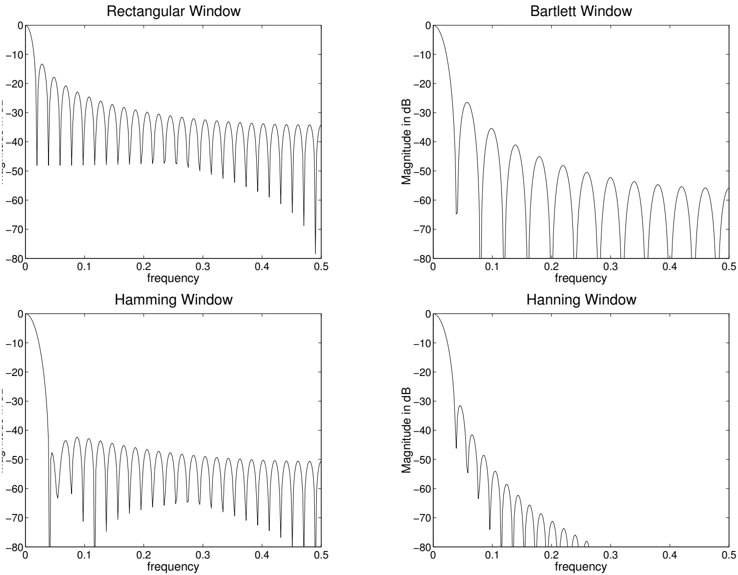
Windows of size $(2M-1)$ samples

The windows satisfy $w(k) \equiv 0$ for $|k| \geq M$, and $w(k) = w(-k)$; the defining equation below are valid for $0 \leq k \leq (M-1)$.

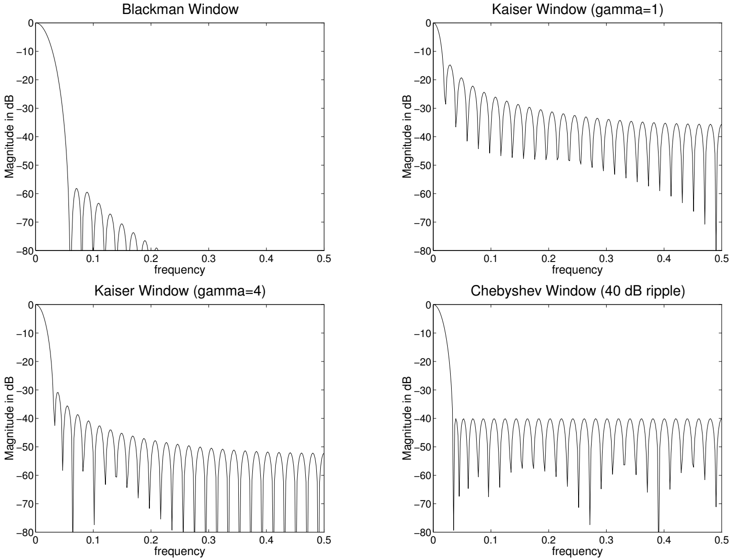
Window Name	Defining Equation	Approx. Main Lobe Width (radians)	Sidelobe Level (dB)
Rectangular	$w(k) = 1$	$2\pi/M$	-13
Bartlett	$w(k) = \frac{M-k}{M}$	$4\pi/M$	-25
Hanning	$w(k) = 0.5 + 0.5\cos\left(\frac{\pi k}{M}\right)$	$4\pi/M$	-31
Hamming	$w(k) = 0.54 + 0.46\cos\left(\frac{\pi k}{M-1}\right)$	$4\pi/M$	-41
Blackman	$w(k) = 0.42 + 0.5\cos\left(\frac{\pi k}{M-1}\right) + 0.08\cos\left(\frac{\pi k}{M-1}\right)$	$6\pi/M$	-57

- For windows of size M samples, multiply main lobe width values by 2.
- Sidelobe levels remain unchanged, as they do not depend on M.

Common windows and their properties (cont'd)



Common windows and their properties (cont'd)



Daniell method: local periodogram smoothing

- Periodogram values $\hat{S}(\omega_k)$ are uncorrelated r.v. at different frequencies ω_k (although strictly speaking this is not usually the case with real data).
- **Idea**: average periodogram around each frequency to lower variance.

Daniell method

$$\hat{S}_{\mathrm{D}}(\omega_k) = \frac{1}{2J+1} \sum_{j=k-J}^{k+J} \hat{S}_{\mathrm{P}}(\omega_j) \qquad \omega_k = \frac{2\pi k}{\tilde{N}}, \quad k = 0, \ldots, \tilde{N}-1.$$

- $\tilde{N}$: DFT length; typically $\tilde{N} \gg N$ for sufficient spectral sampling: zero-padding is needed.
- J chosen such that $S(\omega)$ is nearly constant in $\left[\omega - \frac{2\pi J}{\tilde{N}}, \omega + \frac{2\pi J}{\tilde{N}}\right]$.

Link between Daniell and Blackman-Tukey methods

The Daniell estimator is a Blackman-Tuckey estimator with a rectangular spectral window:

$$W(f) = \begin{cases} 1/\beta & f \in [-\beta/2, \beta/2] \\ 0 & \text{otherwise} \end{cases}$$

where $\beta \stackrel{\text{def}}{=} 2J/\tilde{N}$.

Periodogram averaging

The sample mean: averaging reduces variance (Law of Large Numbers)

Let θ be a random variable with mean μ_θ and variance σ_θ^2. Given K uncorrelated observations $\{\theta_k\}_{k=1}^K$, the **sample mean** is defined as

$$\hat{\mu}_\theta = \frac{1}{K}\sum_{k=1}^{K}\theta_k.$$

Its **mean** and **variance** are given by:

$$\mathrm{E}\{\hat{\mu}_\theta\} = \frac{1}{K}\sum_{k=1}^{K}\mathrm{E}\{\theta_k\} = \mu_\theta$$

$$\mathrm{var}\{\hat{\mu}_\theta\} = \frac{1}{K^2}\sum_{k=1}^{K}\mathrm{var}\{\theta_k\} = \frac{1}{K}\sigma_\theta^2$$

$\rightarrow$ the sample mean is an **unbiased, consistent estimator** of the ensemble mean.

Idea to reduce variance of spectral estimators:

- compute the PSD estimates of several data segments
- compute the sample mean of the PSD estimates.

Bartlett method: periodogram averaging

Bartlett method

1. **Divide** the observed data sequence in K segments of L samples:

$$y_j(n) = y((j-1)L + n), \quad j = 1, 2, \ldots, K, \quad n = 0, 1, \ldots, L-1, \quad \text{with } K \stackrel{\text{def}}{=} \lfloor N/L \rfloor.$$

2. Compute the **periodogram** of each segment:

$$\hat{S}_j(\omega) = \frac{1}{L} \left| \sum_{n=0}^{L-1} y_j(n) e^{-\jmath \omega n} \right|^2.$$

3. **Average** periodograms to produce the Bartlett spectral estimate:

$$\hat{S}_{\mathrm{B}}(\omega) = \frac{1}{K} \sum_{j=1}^{K} \hat{S}_j(\omega).$$

Bartlett method — properties

Mean and variance

$$\mathrm{E}\{\hat{S}_{\mathrm{B}}(\omega)\} = \frac{1}{K}\sum_{j=1}^{K}\mathrm{E}\{\hat{S}_{\mathrm{P}}(\omega)\} = \mathrm{E}\{\hat{S}_{\mathrm{P}}(\omega)\} = S(\omega)*W_{\mathrm{B}}(\omega) \quad \Rightarrow \quad L\text{-sample periodogram}$$

$$\mathrm{var}\{\hat{S}_{\mathrm{B}}(\omega)\} \approx \frac{1}{K^2}\sum_{j=1}^{K}\mathrm{var}\{\hat{S}_{\mathrm{P}}(\omega)\} = \frac{1}{K}S^2(\omega) \quad \Rightarrow \quad \text{variance reduction by factor of } K$$

Comparison with Blackman-Tuckey

- The periodogram of the jth segment can be rewritten as

$$\hat{S}_j(\omega) = \sum_{k=-(L-1)}^{L-1}\hat{r}_j(k)\mathrm{e}^{-\jmath\omega k}.$$

- Hence, Bartlett estimate can be expressed as

$$\hat{S}_{\mathrm{B}}(\omega) = \frac{1}{K}\sum_{j=1}^{K}\sum_{k=-(L-1)}^{L-1}\hat{r}_j(k)\mathrm{e}^{-\jmath\omega k} = \sum_{k=-(L-1)}^{L-1}\frac{1}{K}\sum_{j=1}^{K}\hat{r}_j(k)\mathrm{e}^{-\jmath\omega k} \simeq \sum_{k=-(L-1)}^{L-1}\hat{r}(k)\mathrm{e}^{-\jmath\omega k}$$

$\rightarrow$ $\hat{S}_{\mathrm{B}}(\omega)$ similar to Blackman-Tukey with rectangular window $w_{\mathrm{R}}(k)$ of length $2L-1$.

Welch method: averaged, overlapped, windowed periodograms

Welch method

1. Choose **window size** L and **overlap factor** $\Delta \in [0, 1]$ between consecutive segments.
2. **Divide** the data in K segments of L samples each:

$$y_j(n) = y\big((j-1)D + n\big), \quad j = 1, \ldots, K, \quad n = 0, \ldots, L-1$$

   with $D = \text{round}\big(L(1-\Delta)\big)$ and $K = \lceil (N-L)/D \rceil + 1 \simeq N/D$ if $N \gg L$.
   - no overlap: $\Delta = 0 \quad \Rightarrow \quad D = L \quad \Rightarrow \quad K \simeq N/L$ data segments (as in Bartlett)
   - 50% overlap (typical): $\Delta = 0.5 \quad \Rightarrow \quad D = L/2 \quad \Rightarrow \quad K \simeq 2N/L$ data segments
3. Choose a suitable **window** sequence $w(n)$ of length L.
4. For each segment, compute the **periodogram** of the weighted segment:

$$\hat{S}_j(\omega) = \frac{1}{LU} \left| \sum_{n=0}^{L-1} w(n) y_j(n) \mathrm{e}^{-\jmath\omega t} \right|^2$$

   where $U \stackrel{\text{def}}{=} \frac{1}{L} \sum_{n=0}^{L-1} w(n)^2$ is the power of $w(n)$.
5. **Average** the computed periodograms to produce the Welch estimate:

$$\hat{S}_{\mathrm{W}}(\omega) = \frac{1}{K} \sum_{j=1}^{K} \hat{S}_j(\omega)$$

Welch method — properties

- Decreasing window length L yields more segments and thus reduced variance
  - but if L becomes too small $\rightarrow$ loss of spectral resolution caused by broader main lobe of $W(\omega)$.
- **Overlap** (higher Δ) provides more segments and thus increased averaging and reduced variance, without decreasing window length L
  - but it requires sufficient decorrelation between consecutive segments.
- **Data window**
  - provides mainlobe-sidelobe tradeoff capability to tune compromise between smearing and leakage
  - puts less weight on data samples at the ends of segments $\rightarrow$ reduced crosscorrelation between segments $\rightarrow$ improved variance reduction.
- **Overlap for optimal variance reduction** $\approx 50\%$ of window length ($\Delta = 0.5$).
- $\hat{S}_{\mathrm{W}}(\omega)$ is approximately equal to $\hat{S}_{\mathrm{BT}}(\omega)$ with a non-rectangular lag window:

$$\mathrm{E}\{\hat{S}_{\mathrm{W}}(\omega)\} = S(\omega) * \frac{1}{LU}|W(\omega)|^2$$

$$\mathrm{var}\{\hat{S}_{\mathrm{W}}(\omega)\} \approx \frac{1}{K}\mathrm{var}\{\hat{S}_j(\omega)\}$$

Summary

- Non-parametric estimators can be obtained with computationally efficient operations:
  - DTFT (FFT), convolution, windowing, correlation estimates, ...
- Finite sample size produces biased estimates — two effects:
  - spectral smearing or smoothing (due to window main lobe)
    $\rightarrow$ sets fundamental resolution: $\Delta f_{\text{min}} \sim (\text{window size})^{-1}$
  - power leakage (due to window sidelobes)
- Periodogram is an inconsistent estimator of the PSD $\rightarrow$ variance cannot be reduced by increasing data size (although bias improves)
- To reduce variance, some sort of averaging is necessary:
  - periodogram smoothing: Blackman-Tuckey, Daniell
  - periodogram averaging: Bartlett, Welch.

Summary

- Exploit prior knowledge about the random process to be analyzed
- **General approach**
  1. Select an appropriate model for process under analysis
  2. Estimate the model parameters from the available data
  3. Estimate the PSD by incorporating the estimated parameters into the parametric form for the PSD
- **PSD with rational function structure** — 3 models:
  - autoregressive moving average (ARMA): spectrum with poles and zeros
  - autoregressive (AR): all-pole model
  - moving average (MA): all-zero model
- Model parameters can be estimated using **Yule-Walker equations**
  - estimate ACS $\rightarrow$ build autocorrelation matrix $\rightarrow$ solve a linear system of equations
- Several criteria for **model order selection**
  - AIC, MDL, FPE, ...
- Performance of parametric approach depends on fitness of model to the process being analyzed.