Statistical Signal Processing

Lecture 6

Chapter 2: Spectral Estimation

- review of spectral descriptions of stationary processes
- non-parametric spectral estimation
  - the periodogram
  - spectral leakage, spectral resolution, windowing
  - the averaged periodogram
  - the Blackman-Tukey spectral estimator (the smoothed periodogram)

Stationary Stochastic Processes

- discrete-time stochastic process = sequence of random variables
- description of distribution of stochastic processes often limited to first and second order moments
- for zero mean processes: second-order moments crucial
  in the frequency domain: *power spectral density function* (psdf)
- for non-stationary processes: no ergodicity, no time-averaging. We shall see *time-frequency representations*.
- The random (complex vector) process $\{\mathbf{y}_k\}$ is *wide-sense stationary* (WSS) if its mean (^{H} : *Hermitian* (complex conjugate) transpose)

  $$E\,\mathbf{y}_n = m_\mathbf{y}$$

  does not depend on n and its (matrix) *autocorrelation function* (acf)

  $$r_{\mathbf{yy}}(k) = E\,\mathbf{y}_{n+k}\mathbf{y}_n^H \qquad \text{“acf at lag } k\text{”}$$

  only depends on the time lag between the two samples (not on time). Corresponding central moment : *autocovariance function*

  $$c_{\mathbf{yy}}(k) = E\,(\mathbf{y}_{n+k}-m_\mathbf{y})(\mathbf{y}_n-m_\mathbf{y})^H = r_{\mathbf{yy}}(k) - m_\mathbf{y}\,m_\mathbf{y}^H$$

Stationary Stochastic Processes (2)

- two jointly WSS random vector processes $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ have a *cross-correlation function* (ccf)

$$r_{\mathbf{xy}}(k) \;=\; E\,\mathbf{x}_{n+k}\,\mathbf{y}_n^H$$

and a *cross-covariance function*

$$c_{\mathbf{xy}}(k) \;=\; E\,(\mathbf{x}_{n+k}-m_{\mathbf{x}})(\mathbf{y}_n-m_{\mathbf{y}})^H \;=\; r_{\mathbf{xy}}(k) - m_{\mathbf{x}}\,m_{\mathbf{y}}^H$$

- Unless stated otherwise, we shall consider below processes with zero mean so that the autocorrelation function and the autocovariance function are equal. Nevertheless, the rest of this discussion holds for the general case.

- The following symmetry properties follow immediately from the definition:

$$\begin{aligned} r_{\mathbf{xy}}(k) &= r_{\mathbf{yx}}^H(-k) \\ r_{\mathbf{yy}}(k) &= r_{\mathbf{yy}}^H(-k) \quad . \end{aligned}$$

For real scalar processes

$$\begin{aligned} r_{xy}(k) &= r_{yx}(-k) \\ r_{yy}(k) &= r_{yy}(-k) \quad . \end{aligned}$$

Stationary Stochastic Processes (3)

- In this transparency we consider real scalar processes.

- The following boundedness properties follow from the Cauchy-Schwarz inequality (with correlation as inner product):

$$\begin{aligned} r_{xx}(0)\,r_{yy}(0) &\geq |r_{xy}(k)|^2 \\ r_{yy}(0) &\geq |r_{yy}(k)| \;\geq\; 0 \end{aligned}$$

- The acf is furthermore a *positive semidefinite function*, meaning the following. Take any positive integer M. Let $\{a_j,\, j = 1,\ldots,M\}$ be any set of M real numbers, not all zero, and $\{k_j \in \mathcal{Z},\, j = 1,\ldots,M\}$ any set of M distinct time instants. Then

$$0 \;\leq\; E\,|\sum_{j=1}^{M} a_j\,y_{k_j}|^2 \;=\; \sum_{i=1}^{M}\sum_{j=1}^{M} a_i a_j r_{yy}(k_i-k_j) \;=\; [a_1\cdots a_M]\,T \begin{bmatrix} a_1 \\ \vdots \\ a_M \end{bmatrix}$$

This means also that the symmetric matrix T with elements $T_{ij} = r_{yy}(k_i-k_j)$ is positive semidefinite. T is Toeplitz if the k_j are equidistant.

Spectral Descriptions

- We can consider the z-transforms of the acf and the ccf

$$\begin{aligned} \mathrm{S}_{\mathbf{yy}}(z) &= \sum_{k=-\infty}^{\infty} r_{\mathbf{yy}}(k)\, z^{-k} \\ \mathrm{S}_{\mathbf{xy}}(z) &= \sum_{k=-\infty}^{\infty} r_{\mathbf{xy}}(k)\, z^{-k}\ . \end{aligned}$$

- *auto-psdf* $S_{\mathbf{yy}}(f) = \mathrm{S}_{\mathbf{yy}}(e^{j2\pi f})$ and *cross-psdf* $S_{\mathbf{xy}}(f) = \mathrm{S}_{\mathbf{xy}}(e^{j2\pi f})$

$$\begin{aligned} S_{\mathbf{yy}}(f) &= \sum_{k=-\infty}^{\infty} r_{\mathbf{yy}}(k)\, e^{-j2\pi f k} \qquad \text{Wiener-Khinchin relation} \\ S_{\mathbf{xy}}(f) &= \sum_{k=-\infty}^{\infty} r_{\mathbf{xy}}(k)\, e^{-j2\pi f k}\ . \end{aligned}$$

  auto-psdf and cross-psdf are in principle defined for scalar processes

- Because we normalized the sampling period to 1, the sampling frequency is also normalized to 1 and hence $S_{\mathbf{xy}}(f)$ and $S_{\mathbf{yy}}(f)$ are periodic with period 1.
- We see that estimating $S_{\mathbf{yy}}(f)$ and estimating $r_{\mathbf{yy}}(k)$ are equivalent since both are related by the Fourier transform, a one-to-one transformation.

Spectral Descriptions (2)

- A basic real scalar discrete-time random process is white noise with acf

$$r_{yy}(k) = \sigma_y^2\, \delta_{k0}$$

  where δ_{kn} is the Kronecker delta. This means that the white process is zero mean and all samples are uncorrelated. The psdf becomes

$$S_{yy}(f) = \sigma_y^2$$

  which is constant.

- LTI filter transfer fn: $\boldsymbol{H}(f) = \sum_m \mathbf{h}_m\, e^{-j2\pi f m} = \mathbf{H}(e^{j2\pi f})$, $\mathbf{H}(z) = \sum_m \mathbf{h}_m\, z^{-m}$
- The filtering of a WSS process by a linear time-invariant (LTI) filter produces another WSS process. Indeed, let $\mathbf{y}_n = \sum_m \mathbf{h}_{n-m}\mathbf{x}_m$ where $\{\mathbf{x}_k\}$ is WSS. Then we get for the mean

$$E\,\mathbf{y}_n = \sum_m \mathbf{h}_m\, E\,\mathbf{x}_{n-m} = m_{\mathbf{x}} \sum_m \mathbf{h}_m = \boldsymbol{H}(0)\, m_{\mathbf{x}} = m_{\mathbf{y}}$$

  which does indeed not depend on n. Furthermore,

$$\begin{aligned} E\,\mathbf{y}_{k+n}\mathbf{x}_n^H &= \sum_m \mathbf{h}_{k+n-m}\, E\,\mathbf{x}_m\mathbf{x}_n^H = \sum_m \mathbf{h}_{k+n-m}\, r_{\mathbf{xx}}(m-n) = \sum_m \mathbf{h}_{k-m}\, r_{\mathbf{xx}}(m) \\ &= \mathbf{h}_k * r_{\mathbf{xx}}(k) = r_{\mathbf{yx}}(k) \end{aligned}$$

Spectral Descriptions (3)

and similarly

$$\begin{aligned} E\,\mathbf{x}_{k+n}\mathbf{y}_n^H &= \sum_m (E\,\mathbf{x}_{k+n}\,\mathbf{x}_m^H)\,\mathbf{h}_{n-m}^H \;=\; \sum_m r_{\mathbf{xx}}(k{+}n{-}m)\,\mathbf{h}_{n-m}^H \;=\; \sum_m r_{\mathbf{xx}}(k{-}m)\,\mathbf{h}_{-m}^H \\ &= r_{\mathbf{xx}}(k) * \mathbf{h}_{-k}^H \;=\; r_{\mathbf{xy}}(k) \end{aligned}$$

$$\begin{aligned} E\,\mathbf{y}_{k+n}\mathbf{y}_n^H &= \sum_m \mathbf{h}_{k+n-m} \sum_l (E\,\mathbf{x}_m\,\mathbf{x}_l^H)\,\mathbf{h}_{n-l}^H \;=\; \sum_m \mathbf{h}_{k+n-m} \sum_l r_{\mathbf{xx}}(m{-}l)\,\mathbf{h}_{n-l}^H \\ &= \sum_m \mathbf{h}_{k-m} \sum_l r_{\mathbf{xx}}(m{-}l)\,\mathbf{h}_{-l}^H \;=\; \sum_m \mathbf{h}_{k-m} r_{\mathbf{xy}}(m) \;=\; \mathbf{h}_k * r_{\mathbf{xy}}(k) \\ &= \mathbf{h}_k * r_{\mathbf{xx}}(k) * \mathbf{h}_{-k}^H \;=\; r_{\mathbf{yy}}(k) \end{aligned}$$

in which the correlation considered is each time independent of time. This shows that not only $\{\mathbf{y}_k\}$ is WSS but furthermore $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ are jointly WSS.

- Taking z- and Fourier transforms of these convolution relations yields

$$\begin{aligned} \mathrm{S}_{\mathbf{yx}}(z) &= \mathbf{H}(z)\,\mathrm{S}_{\mathbf{xx}}(z) & S_{\mathbf{yx}}(f) &= \boldsymbol{H}(f)\,S_{\mathbf{xx}}(f) \\ \mathrm{S}_{\mathbf{xy}}(z) &= \mathrm{S}_{\mathbf{xx}}(z)\,\mathbf{H}^{\dagger}(z) & S_{\mathbf{xy}}(f) &= S_{\mathbf{xx}}(f)\,\boldsymbol{H}^H(f) \\ \mathrm{S}_{\mathbf{yy}}(z) &= \mathbf{H}(z)\,\mathrm{S}_{\mathbf{xx}}(z)\,\mathbf{H}^{\dagger}(z) & S_{\mathbf{yy}}(f) &= \boldsymbol{H}(f)\,S_{\mathbf{xx}}(f)\,\boldsymbol{H}^H(f) \end{aligned}$$

w. $\mathbf{H}(z) = \sum_k \mathbf{h}_k z^{-k}$ and for real h_k: $S_{yy}(f) = |H(f)|^2 S_{xx}(f)$.
paraconjugate (*matched* filter): $\mathbf{H}^{\dagger}(z) = \mathbf{H}^H(1/z^*)$ = z-transform of $\mathbf{h}_{-k}^H$

Power Spectral Density Interpretation

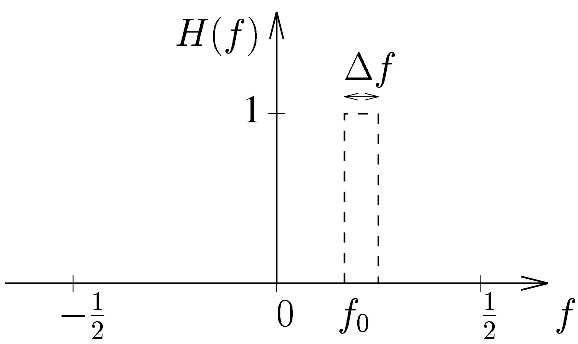
- The last relation leads to the interpretation of $S_{xx}(f)$ as power spectral density function. Indeed, on the one hand we find from the Fourier relationship

$$r_{xx}(0) \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{xx}(f)df$$

which means that the total power $r_{xx}(0) = E\,x^2(k)$ is made up of the sum of the contributions at all frequencies. *real scalar process considered here*

- Now let $\{x_k\}$ and $\{y_k\}$ be the input-output pair associated with a LTI filter with transfer function

$$H(f) \;=\; \begin{cases} 1 \;, & f \in [f_0, f_0+\Delta f]\,,\ \Delta f \text{ arbitrarily small} \\ 0 \;, & \text{elsewhere in } [-\frac{1}{2}, \frac{1}{2}] \end{cases}$$



Power Spectral Density Interpretation (2)

- Assume that $S_{xx}(f)$ is continuous at f_0. We get

$$r_{yy}(0) \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{yy}(f)df \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f)|^2\, S_{xx}(f)df \;=\; \int_{f_0}^{f_0+\Delta f} S_{xx}(f)df \;=\; S_{xx}(f_0)\,\Delta f$$

  or $S_{xx}(f_0) \;=\; r_{yy}(0)/\Delta f$ which first of all implies $S_{xx}(f) \;\geq\; 0$ and secondly justifies the name power spectral density of $S_{xx}(f)$.

- For the special case of $\{x_k\}$ being a white noise input, the psdf of the output process $\{y_k\}$ of a LTI filter with transfer function $H(f)$ becomes

$$S_{yy}(f) \;=\; \sigma_x^2\, |H(f)|^2$$

- We have shown independently that the acf is a positive semidefinite function and that the psdf is nonnegative. It is possible to show that these two properties imply each other. Indeed, if $r_{xx}(k)$ and $S_{xx}(f)$ are two functions that form a Fourier transform pair, then the following theorem holds.

  Theorem (Bochner's theorem) $r_{xx}(k)$ *is real, even and positive (semi)definite iff* $S_{xx}(f)$ *is real, even and positive (nonnegative).*

Fourier Transform Correlation

- the Fourier transform $\boldsymbol{Y}(f) \;=\; \sum_k \mathbf{y}_k e^{-j2\pi fk}$ of the stationary process $\{\mathbf{y}_k\}$ is random

- **Theorem (Fourier transform correlation)**
  Let $\boldsymbol{X}(f) = \sum\limits_{k=-\infty}^{\infty} \mathbf{x}_k e^{-j2\pi fk}$ *and* $\boldsymbol{Y}(f) = \sum\limits_{k=-\infty}^{\infty} \mathbf{y}_k e^{-j2\pi fk}$. *Then*

$$E\,\boldsymbol{X}(f)\boldsymbol{Y}^H(f_1) \;=\; S_{\mathbf{xy}}(f)\,\delta_1(f-f_1)$$

  where we have introduced the impulse train $\delta_{f_0}(f) \;=\; \sum\limits_{n=-\infty}^{\infty} \delta(f-nf_0)$.

- In particular, $E\,Y(f)Y^*(f_1) \;=\; S_{yy}(f)\,\delta_1(f-f_1)$ which means that a real scalar random process $y_k \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} Y(f)\,e^{j2\pi fk}df$ can be regarded as a superposition of exponentials $e^{j2\pi fk}$ with random complex amplitudes $Y(f)$ that are uncorrelated at different frequencies and that have a power proportional to $S_{yy}(f)$.

Filtering Relationships

- Using the above result, it is straightforward to show the effect on the auto- or cross-psdf of linear filtering. Indeed consider the processes $\mathbf{u}_k$, $\mathbf{v}_k$, $\mathbf{x}_k$ and $\mathbf{y}_k$ which are related by linear filtering as shown in the figure. Then we get

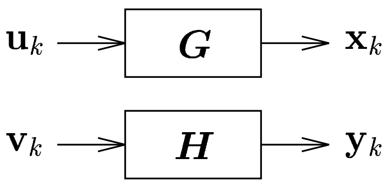
$$\begin{aligned} S_{\mathbf{xy}}(f)\delta_1(f-f_1) &= E\,\boldsymbol{X}(f)\boldsymbol{Y}^H(f_1) \;=\; \boldsymbol{G}(f)\left(E\,\boldsymbol{U}(f)\boldsymbol{V}^H(f_1)\right)\boldsymbol{H}^H(f_1) \\ &= \boldsymbol{G}(f)S_{\mathbf{uv}}(f)\boldsymbol{H}^H(f_1)\delta_1(f-f_1)\;. \end{aligned}$$

  Integrating both sides, $\int_{-\frac{1}{2}}^{\frac{1}{2}} df_1$, yields

$$S_{\mathbf{xy}}(f) \;=\; \boldsymbol{G}(f)S_{\mathbf{uv}}(f)\boldsymbol{H}^H(f)\;.$$

- We can extend this property to the z-transforms:

$$\mathrm{S}_{\mathbf{xy}}(z) \;=\; \mathbf{G}(z)\mathrm{S}_{\mathbf{uv}}(z)\mathbf{H}^{\dagger}(z)$$



$$\begin{aligned} \boldsymbol{X}(f) &= \boldsymbol{G}(f)\,\boldsymbol{U}(f) \\ \boldsymbol{Y}(f) &= \boldsymbol{H}(f)\,\boldsymbol{V}(f) \end{aligned}$$

Why Spectral Estimation

- The spectrum transmitted by modems needs to satisfy certain restrictions, especially if the modem needs to comply with a certain standard. For instance voiceband modems over the telephone line need to be restricted to the [300Hz,3400Hz] band. Wireless modems or any radio transmitter needs to occupy a rectricted bandwidth since the radio medium is scarce. In these considerations, not only the (effective) bandwidth of the signal is of importance but also the spectral roll-off. Typically, the emitted spectrum needs to fall under a certain spectral mask (plot of the maximum allowed power spectral density as a function of frequency). To verify this, the spectrum of the emitted signal needs to be estimated.

- We shall see that the optimization of the parameters of most source coding techniques depends on the spectrum or equivalently the correlation structure of the source to be coded. Hence, spectral estimation is again required. Such sources can be speech, audio, images or video.

- Spectrum estimation also allows to evaluate certain parameters about transmitted signals such as the frequency offset and the frequency-selective distortion introduced by the channel, or the spectrum of the interfering signals or noise.

Spectral Estimation

- There exist two big methodologies for estimating the psdf of a stationary process: *non-parametric* techniques and *parametric* techniques.

- In non-parametric or *classical* spectral estimation techniques, no constraints are imposed on the possible form of $S_{yy}(f)$. This implies that an infinite number of degrees of freedom need to be estimated: the $r_{yy}(k)$. This will lead to a very high estimation variance if we have to work with a finite amount of data.

- In the parametric or *high-resolution* techniques, a parametric model is assumed for $S_{yy}(f)$ and the spectral estimatiom problem reduces to the estimation of the parameters describing the parametric model. Since the parametric form can only perfectly model a limited class of functions, there will be some bias in the estimate of $S_{yy}(f)$. However, to compensate for this bias, a (large) reduction in variance is achievable compared to the non-parametric techniques.

Non-Parametric Spectral Estimation

- The first non-parametric technique is based on the Fourier transform. Consider the Fourier transform $Y(f)$ of the WSS process y_k with zero mean. $Y(f)$ is a random function with mean

$$E\,Y(f) \;=\; \sum_{k=-\infty}^{\infty} e^{-j2\pi fk}\, \underbrace{E\,y_k}_{=0} \;=\; 0\;.$$

The correlation between the Fourier transform at frequencies f and f_1 is

$$E\,Y(f)Y^*(f_1) \;=\; S_{yy}(f)\delta_1(f-f_1)\;.$$

So the Fourier transform at different frequencies is uncorrelated, while its magnitude squared is proportional to the power spectral density function.

- In practice, we are given only a finite number of N samples $\{y_0, y_1, \ldots, y_{N-1}\}$. The *periodogram* was introduced by Schuster in 1898 and is defined as the scaled magnitude squared of the Fourier transform of this finite number of samples:

$$\widehat{S}_{yy}(f) \;=\; \widehat{S}_{PER}(f) \;=\; \frac{1}{N}\left|\sum_{n=0}^{N-1} y_n\, e^{-j2\pi fn}\right|^2$$

where $\frac{1}{N}$ has been introduced to avoid that things $\to\infty$ as $N\to\infty$.

The Periodogram

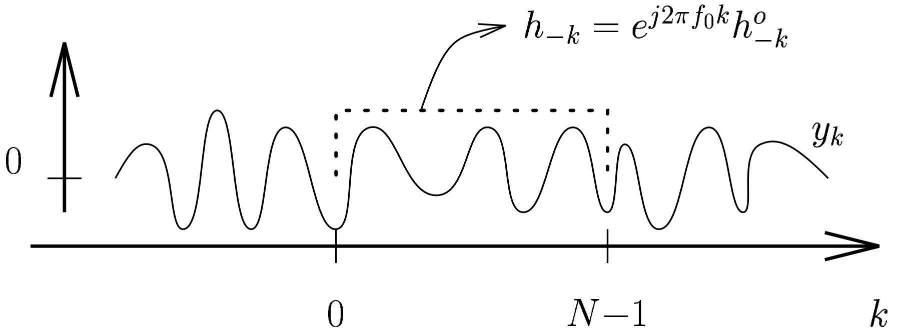
- focus on a particular frequency: denote f as f_0. Interpretation:

$$\widehat{S}_{PER}(f_0) \;=\; \frac{1}{N}\left|\sum_{n=0}^{N-1} y_n\, e^{-j2\pi f_0 n}\right|^2 \;=\; N\left|\sum_{n=0}^{N-1} h^o_{k-n} y_n\right|^2_{k=0}$$

where $h^o_k = h_k\, e^{j2\pi f_0 k}$ and

$$h_k \;=\; \begin{cases} \frac{1}{N} & ,\ k = -(N-1), \ldots, -1, 0 \\ 0 & ,\ \text{otherwise} \end{cases}$$

$h^o_k \;=\;$ anticausal FIR filter = a modulated rectangular window h_k



The Periodogram (2)

- The frequency response of the filter h_k is

$$H(f) \;=\; \sum_{k=-\infty}^{\infty} h_k\, e^{-j2\pi f k} \;=\; \frac{1}{N}\sum_{k=-(N-1)}^{0} e^{-j2\pi f k} \;=\; \frac{1}{N}\sum_{k=0}^{N-1} e^{j2\pi f k}$$

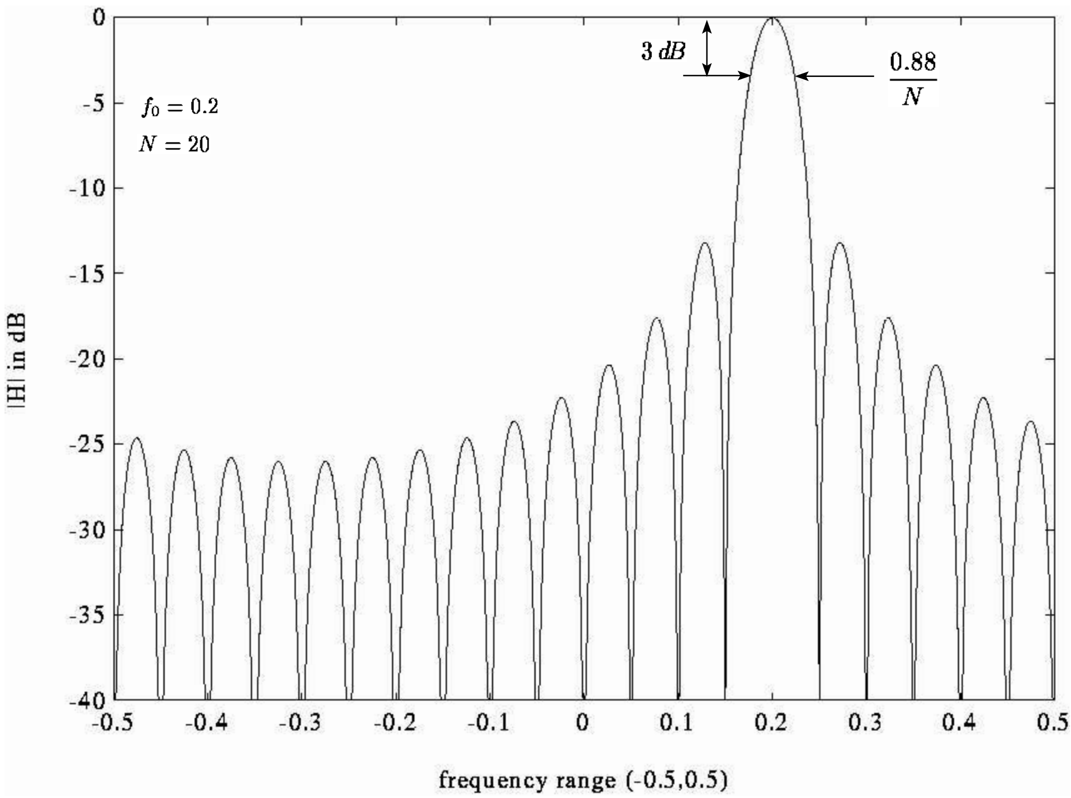
$$= \frac{1}{N}\,\frac{1-e^{j2\pi f N}}{1-e^{j2\pi f}} \;=\; \frac{1}{N}\,\frac{e^{-j\pi f N}-e^{j\pi f N}}{e^{-j\pi f}-e^{j\pi f}}\,\frac{e^{j\pi f N}}{e^{j\pi f}} \;=\; \frac{\sin N\pi f}{N\,\sin \pi f}\, e^{j(N-1)\pi f}$$

- $H(f)$ is of course periodic with period 1. It satisfies furthermore

$$\begin{array}{l} H(k) = 1 \ \ ,\ k \in \mathcal{Z} \\ H(\frac{k}{N}) = 0 \ ,\ k \in \mathcal{Z} \setminus N\mathcal{Z} \ . \end{array}$$

- $H^o(f) \;=\; H(f{-}f_0)$ is the frequency response of a *bandpass filter* with center frequency f_0. The 3dB bandwidth is $\frac{0.88}{N}$ $\;\Rightarrow\;$ bandwidth $\approx \frac{1}{N}$.
- Hence the periodogram estimates the power in $\{y_k\}$ at the frequency f_0 by filtering the data with a bandpass filter, sampling the output at time $k = 0$, and computing the magnitude squared. When multiplied by N $(= \frac{1}{\Delta f} = \frac{1}{1/N})$ (to account for the bandwidth of the bandpass filter), this yields the power spectral density estimate $\widehat{S}_{PER}(f_0)$ (compare to psdf interpretation).

The Periodogram (3)



The Periodogram (4)

- If $N \to \infty$, then $\widehat{S}_{PER}(f_0) \to S_{yy}(f_0)$?
  Is the Periodogram a consistent estimator of the psdf?

- White noise example: it appears that the random fluctuations or variance of the periodogram does not decrease with N, $\Rightarrow$ the periodogram is *not* a consistent estimator of the psdf. Nevertheless, for this white noise example, the periodogram appears to fluctuate around a constant value (the true constant value of the psdf), $\Rightarrow$ periodogram unbiased for white noise. But the variance does not tend to zero as $N \to \infty$.

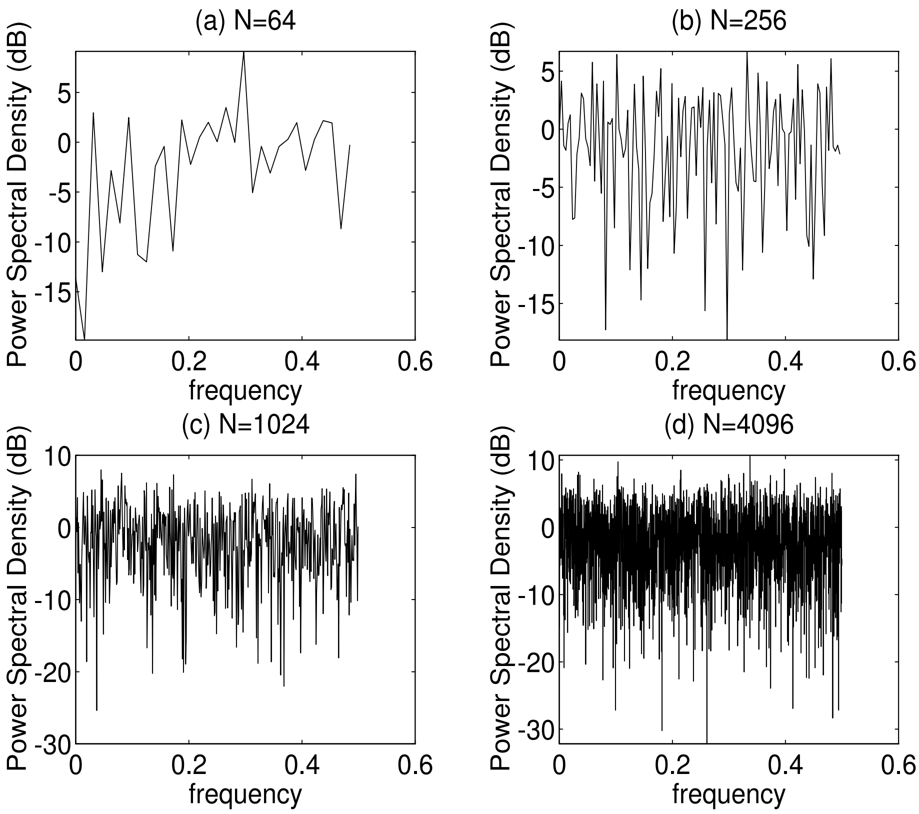
- Intuitively, in order to have a consistent estimator of a set of parameters, we need to have lots more data than the number of parameters to be estimated. The number of parameters is ∞ here: Wiener-Khinchin relation

$$S_{yy}(f) \;=\; r_{yy}(0) + 2\sum_{n=1}^{\infty} r_{yy}(n)\,\cos(2\pi f n)$$

  $\{r_{yy}(k),\, k \geq 0\}$ = infinite set of parameters parameterizing $S_{yy}(f)$. Even if $N \to \infty$, $N \gg$ the number of parameters is impossible. Hence, the variance in estimating those parameters cannot go to zero.

The Periodogram (5)

Illustrating the periodogram inconsistency for white Gaussian noise ($\sigma_y^2 = 1$) $10\log_{10}\widehat{S}_{PER}(f)$ for $N = 64$ (a), 256 (b), 1024 (c), 4096 (d).



Periodogram Mean

- We get for the convolution of h_k^o and y_k (inverse Fourier T of its FT)

$$\sum_n h_{k-n}^o y_n = h_k^o * y_k = \mathcal{F}^{-1}\{H^o(f)Y(f)\}$$

where h_k^o and $H^o(f) = H(f-f_0)$ depend on f_0. Hence

$$\left[\sum_{n=0}^{N-1} h_{k-n}^o y_n\right]_{k=0} = \left[\int_{-\frac{1}{2}}^{\frac{1}{2}} H^o(f)Y(f)e^{j2\pi fk}df\right]_{k=0} = \int_{-\frac{1}{2}}^{\frac{1}{2}} H^o(f)Y(f)df\ .$$

- So we get for the mean of the periodogram

$$\begin{aligned}
E\,\widehat{S}_{PER}(f_0) &= N\,E\left|\left[\Sigma_{n=0}^{N-1} h_{k-n}^o y_n\right]_{k=0}\right|^2 = N\,E\left|\int_{-\frac{1}{2}}^{\frac{1}{2}} H^o(f)Y(f)df\right|^2 \\
&= N\,E\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} H^o(f)Y(f)df\right)\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} H^{o*}(f_1)Y^*(f_1)df_1\right) \\
&= N\int_{-\frac{1}{2}}^{\frac{1}{2}} df\, H^o(f)\int_{-\frac{1}{2}}^{\frac{1}{2}} df_1 H^{o*}(f_1)\,\underbrace{E\,Y(f)Y^*(f_1)}_{=S_{yy}(f)\delta_1(f-f_1)} \\
&= N\int_{-\frac{1}{2}}^{\frac{1}{2}} df\, H^o(f)S_{yy}(f)\,\underbrace{\int_{-\frac{1}{2}}^{\frac{1}{2}} df_1 H^{o*}(f_1)\delta_1(f-f_1)}_{=H^{o*}(f)} = N\int_{-\frac{1}{2}}^{\frac{1}{2}} |H^o(f)|^2\,S_{yy}(f)df \\
&= N\int_{-\frac{1}{2}}^{\frac{1}{2}} |H(f-f_0)|^2\,S_{yy}(f)df = \int_{-\frac{1}{2}}^{\frac{1}{2}} \underbrace{N\,|H(f_0-f)|^2}_{=W_B(f_0-f)} S_{yy}(f)df
\end{aligned}$$

Periodogram Mean (2)

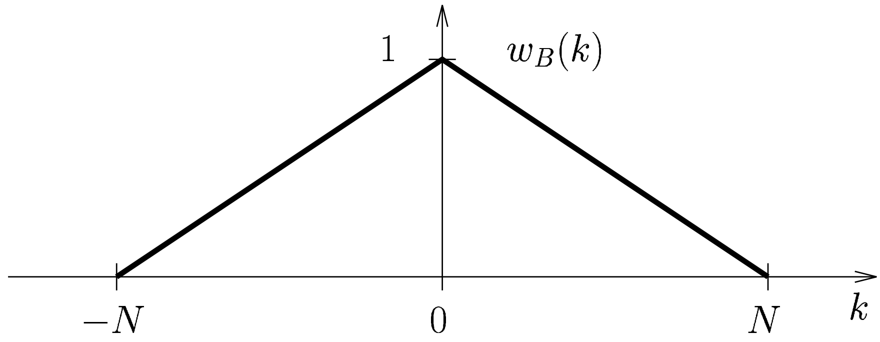
- Hence, $E\,\widehat{S}_{PER}(f_0) \;=\; \int_{-\frac{1}{2}}^{\frac{1}{2}} W_B(f_0-f)S_{yy}(f)df \;=\; W_B(f_0) * S_{yy}(f_0)$ where

$$W_B(f) \;=\; \frac{1}{N}\left(\frac{\sin \pi f N}{\sin \pi f}\right)^2 \;=\; N\,|H(f)|^2\,.$$

- Since $W_B(f)$ is (apart from the scaling factor N) the square of the Fourier transform of a rectangular window, it is also the Fourier transform of the convolution of a rectangular window with itself which is a triangular window. Hence

$$w_B(k) \;=\; \mathcal{F}^{-1}\left\{W_B(f)\right\} \;=\; \begin{cases} 1-\frac{|k|}{N} & ,\ |k| \le N-1 \\ 0 & ,\ |k| \ge N\,. \end{cases}$$

Such a window is called a *Bartlett window*.



Periodogram Mean (3)

- We conclude that the average periodogram is the convolution of the true psdf with the Fourier transform of the Bartlett window, yielding on the average a smoothed version of the psdf. Now since in general $W_B(f) * S_{yy}(f) \neq S_{yy}(f)$, the periodogram is biased in general for finite data records.

- However, the periodogram is asymptotically unbiased. Indeed

$$\begin{aligned}\lim_{N\to\infty} E\,\widehat{S}_{PER}(f) \;&=\; \lim_{N\to\infty} W_B(f) * S_{yy}(f) \;=\; \lim_{N\to\infty} \mathcal{F}\left\{w_B(k) r_{yy}(k)\right\} \\ &=\; \mathcal{F}\{\underbrace{\lim_{N\to\infty} w_B(k)}_{=1}\; r_{yy}(k)\} \;=\; \mathcal{F}\left\{r_{yy}(k)\right\} \;=\; S_{yy}(f)\,.\end{aligned}$$

- Note that

$$w_B(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} W_B(f)df = 1\,,\quad W_B(0) = N\,,\quad W_B(f) \ge 0\,,$$
$$6dB \text{ bandwidth of } W_B(f) \;=\; \tfrac{0.88}{N}\,.$$

As $N \to \infty$, $W_B(f)$ becomes a narrow and high peak of constant surface. So $W_B(f) \to \delta(f)$. For the special case of white noise, the periodogram is unbiased even for finite data records.

Periodogram Covariance

- As far as the covariance of the periodogram is concerned, the following can be shown exactly for white Gaussian noise. The same result holds true approximately for more general processes when the data record is large.

$$Cov\left[\widehat{S}_{PER}(f_1),\widehat{S}_{PER}(f_2)\right] \approx S_{yy}(f_1)S_{yy}(f_2)\left[W_B(f_1+f_2)+W_B(f_1-f_2)\right]\frac{1}{N}$$

- The variance at the frequency f then follows as

$$Var\left[\widehat{S}_{PER}(f)\right] \;=\; Cov\left[\widehat{S}_{PER}(f),\widehat{S}_{PER}(f)\right] \approx S^2_{yy}(f)\left[1+\frac{W_B(2f)}{N}\right] \;\geq\; S^2_{yy}(f)\ .$$

- For frequencies not near 0 or $\pm\frac{1}{2}$, we can further approximate as

$$Var\left[\widehat{S}_{PER}(f)\right] \approx S^2_{yy}(f)$$

  which is independent of the record length N!

- Furthermore, modulo certain conditions, the following result holds exactly

$$\lim_{N\to\infty} Var\left[\widehat{S}_{PER}(f)\right] \;=\; \begin{cases} 2\,S^2_{yy}(f) & ,\quad f=0,\frac{1}{2} \\ S^2_{yy}(f) & ,\quad f\in(0,\frac{1}{2})\ . \end{cases}$$

Periodogram Covariance (2)

- Hence the periodogram is an unreliable estimator since its standard deviation is approximately as large as the (nonnegative) quantity to be estimated.

- We get also

$$Cov\left[\widehat{S}_{PER}(f_1),\widehat{S}_{PER}(f_2)\right] \approx 0 \text{ if } f_1 \neq f_2 \text{ are integer multiples of } \tfrac{1}{N}.$$

  The values of the periodogram at integer multiples of $\frac{1}{N}$ are uncorrelated.

- Coupled with the constant variance (as a function of N), this implies that as N increases the periodogram fluctuates rapidly as illustrated for white noise.

- In fact, it can also be shown that

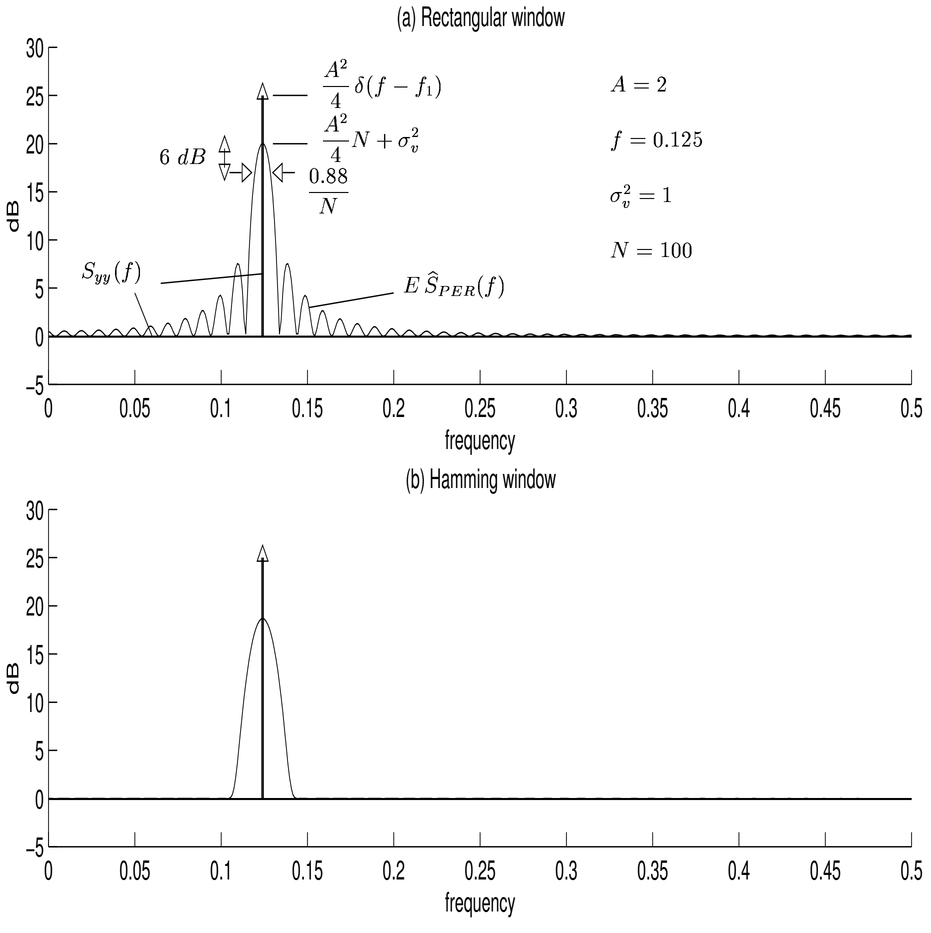
$$\lim_{N\to\infty} Cov\left[\widehat{S}_{PER}(f_1),\widehat{S}_{PER}(f_2)\right] \;=\; 0\ ,\quad f_1\neq f_2$$

  so that asymptotically, the periodogram is uncorrelated at different frequencies, just like the Fourier transform of the whole realization y_k.

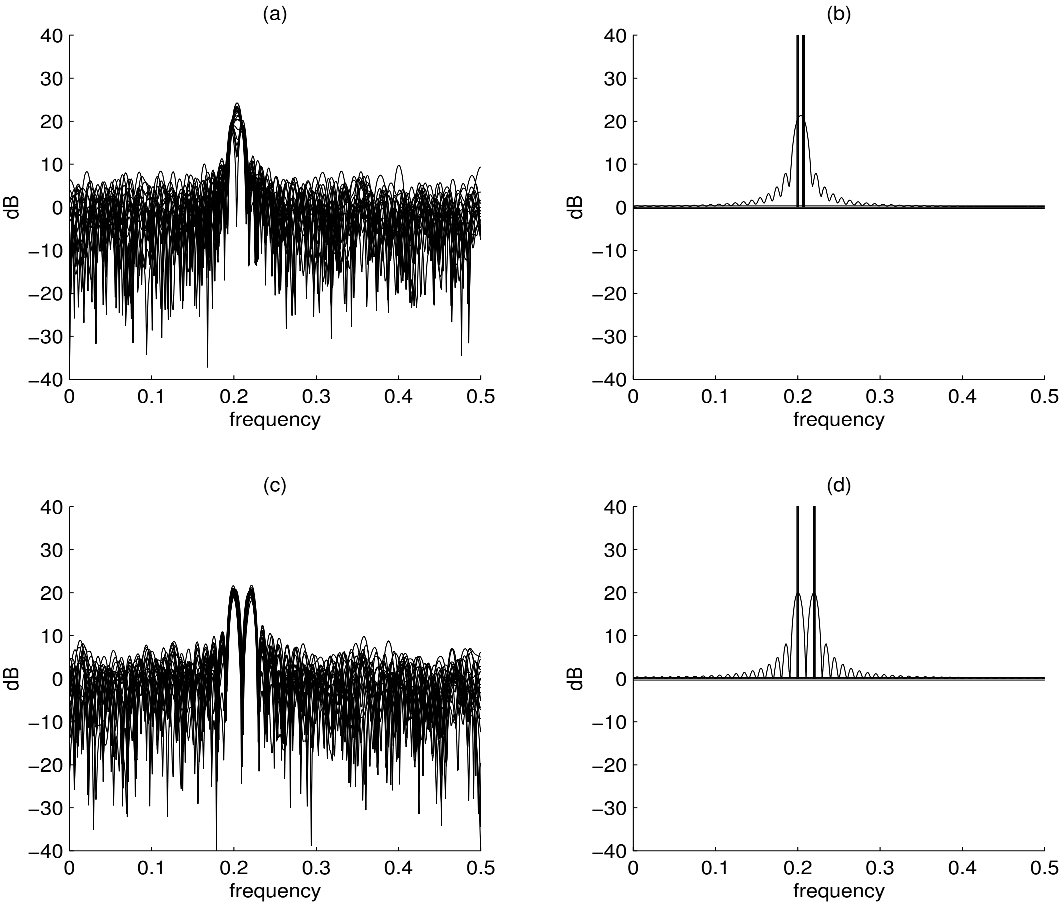
Spectral Leakage

- If the process y_k under consideration would be a (complex) sinusoid with frequency f_0, $y_k = e^{j(2\pi f_0 k + \phi)}$ with ϕ uniformly distributed over $[0, 2\pi)$, then $r_{yy}(n) = e^{j2\pi f_0 n}$, $S_{yy}(f) = \delta(f - f_0)$
  and the periodogram mean would be $W_B(f - f_0) = N\,|H(f - f_0)|^2$.
- strong sidelobes $\Rightarrow$ the average periodogram has non-zero contributions at frequencies where the psdf has no contributions: *spectral leakage*.
- to reduce this spectral leakage, replace the rectangular window with another window that shows less discontinuities near the edges and hence that has weaker sidelobes.
- The price to pay is that the bandwidth of the main lobe increases, reducing the *resolution*. For maximum resolution, or the ability to observe two comparable level sinusoids, no data windowing should be used. Considering the 3 dB bandwidth of $|H(f)|$, a common rule of thumb is that two equiamplitude sinusoids are resolvable if their normalized frequencies are spaced more than $1/N$ apart. If the data consists of only one sinusoid (or several sinusoids spaced much more than $1/N$ apart) embedded in white Gaussian noise, then the optimal (ML) frequency estimator is the periodogram without data windowing.

Spectral Leakage (2)



Illustrating Spectral Resolution



$(a), (b)$:
$(f_1, f_2) = (2, 2.07)$

$(c), (d)$:
$(f_1, f_2) = (2, 2.2)$

$N = 100$
$A_1 = A_2 = 2$
$\sigma_v^2 = 1$
rectang. wind.

Data Weighting

- Since the windows used in practice are symmetric w.r.t. the center of the data record, we shall assume for this windowing discussion that $N = 2M+1$ and that the available data are $y_{-M}, \ldots, y_M$. So we apply a symmetric window $w_k = w_{-k}$ before computing the periodogram. The periodogram with weighting becomes:

$$\widehat{S}_{PER,w}(f) \;=\; \frac{1}{2M+1}\left|\sum_{k=-M}^{M} w_k y_k e^{-j2\pi f k}\right|^2 .$$

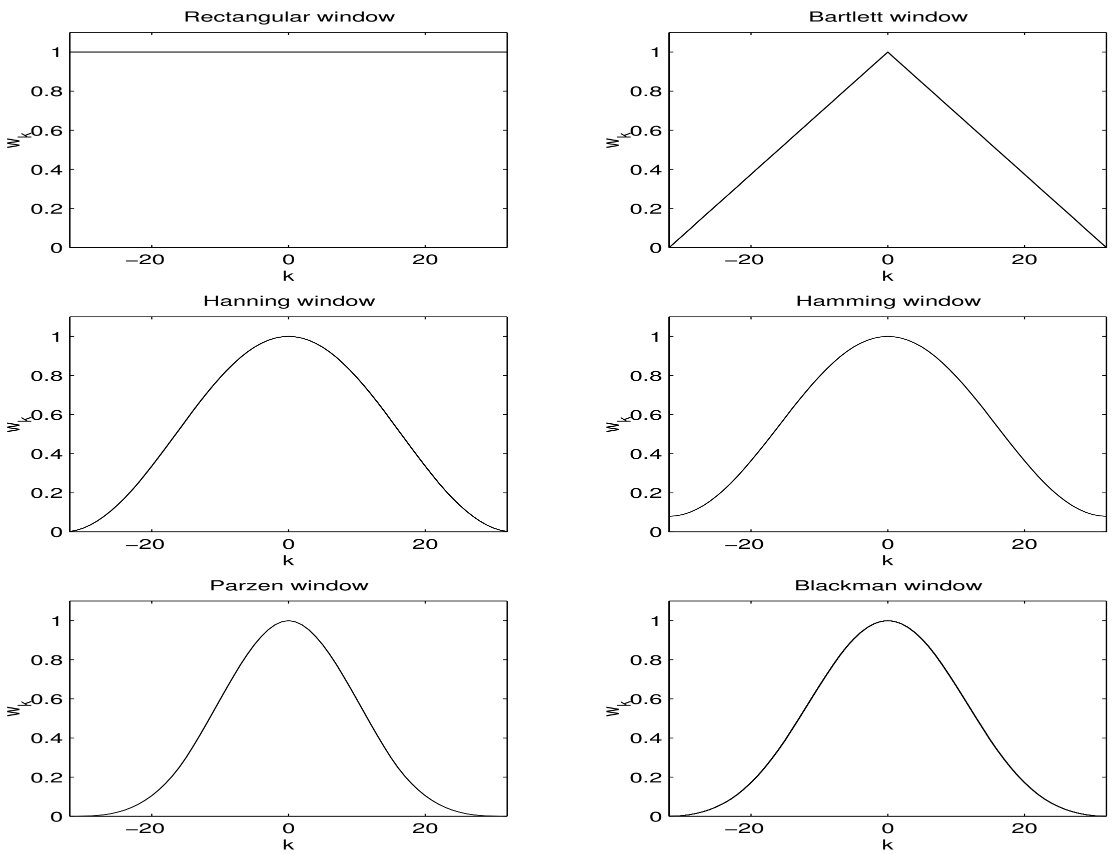
Some popular windows are (all with $w_k = 0,\; |k| > M$):

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

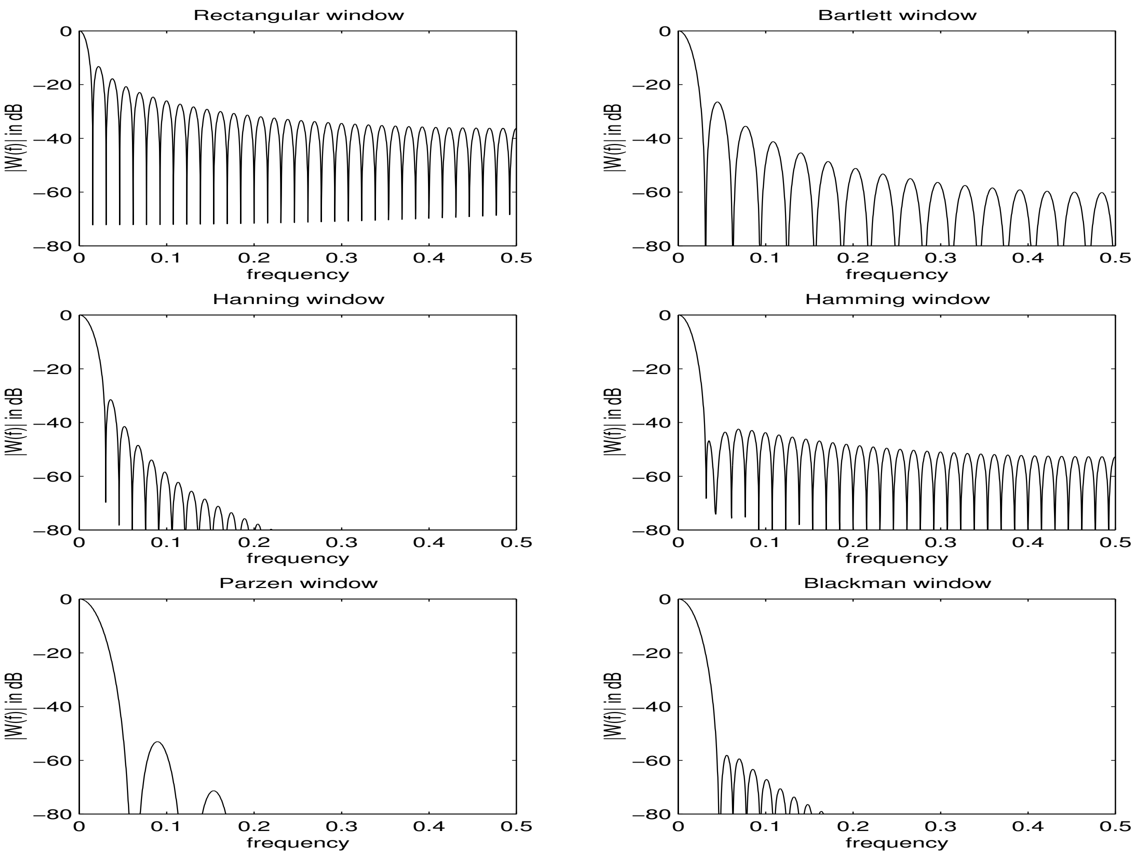
Data Windowing

Name	Definition: $w_k =$	Fourier transform: $W(f) =$	$\mathcal{B}_{3dB}$	A_{SL}
Rectangular	1	$W_R(f) = \frac{\sin \pi f(2M+1)}{\sin \pi f}$	$\frac{0.88}{N}$	-13
Bartlett	$1-\frac{\lvert k\rvert}{M}$	$W_B(f) = \frac{1}{M}\left(\frac{\sin \pi f M}{\sin \pi f}\right)^2$	$\frac{1.28}{N}$	-27
Hanning	$\frac{1}{2}+\frac{1}{2}\cos\frac{\pi k}{M}$	$\frac{1}{2}W_R(f)+$ $\frac{1}{4}(W_R(f-\frac{1}{2M})+W_R(f+\frac{1}{2M}))$	$\frac{1.44}{N}$	-32
Hamming	$0.54+0.46\cos\frac{\pi k}{M}$	$0.54W_R(f)+$ $0.23(W_R(f-\frac{1}{2M})+W_R(f+\frac{1}{2M}))$	$\frac{1.30}{N}$	-43
Parzen (M even)	$\begin{cases} 2(1-\frac{\lvert k\rvert}{M})^3-(1-2\frac{\lvert k\rvert}{M})^3, & \lvert k\rvert \le \frac{M}{2} \\ 2(1-\frac{\lvert k\rvert}{M})^3, & \frac{M}{2} < \lvert k\rvert \le M \end{cases}$	$\frac{8}{M^3}\left(\frac{3}{2}\frac{\sin^4 \pi f M/2}{\sin^4 \pi f} - \frac{\sin^4 \pi f M/2}{\sin^2 \pi f}\right)$	$\frac{1.84}{N}$	-53
Blackman	$0.42+0.5\cos\frac{\pi k}{M}+0.08\cos\frac{2\pi k}{M}$	$0.42W_R(f)+$ $0.25(W_R(f-\frac{1}{2M})+W_R(f+\frac{1}{2M}))$ $+0.04(W_R(f-\frac{1}{M})+W_R(f+\frac{1}{M}))$	$\frac{1.68}{N}$	-58

Data Windows in Time



Data Windows in Frequency

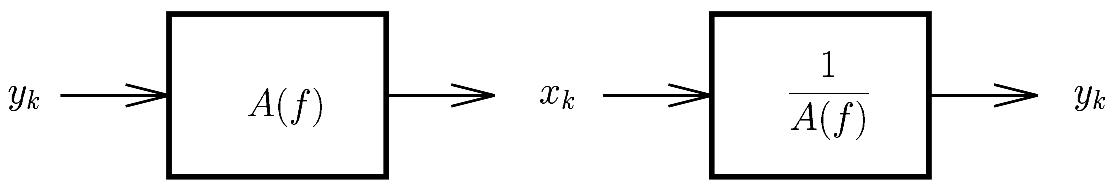


Bias Reduction by Data Prewhitening

- If the psdf is constant, then the periodogram is unbiased, even for finite data record length N.

- $\Rightarrow$ idea for bias reduction: Suppose we have some a priori idea about the psdf (or perhaps, after a first application of the periodogram, we have such information). Then we may design a filter $A(f)$ such that the result x_k obtained by filtering y_k with $A(f)$ is approximately white (the technique of linear prediction that we shall see further provides one convenient way of doing this). Then we determine the periodogram $\widehat{S}_{xx,PER}(f)$ of the filtered data. The bias in $\widehat{S}_{xx,PER}(f)$ will be small since x_k is approximately white. The psdf estimate for the original data y_k can then be taken to be

$$\widehat{S}_{yy}(f) = \frac{\widehat{S}_{xx,PER}(f)}{|A(f)|^2}.$$



Use of the DFT and Zero Padding

- The computation of the periodogram: evaluate the Fourier transform of the windowed data for continuous f. The computer can only handle numbers and hence a discrete set of frequencies $\Rightarrow$ Discrete Fourier Transform (DFT) (the Fast Fourier Transform (FFT)). For consideration of the DFT/FFT, it is customary to consider the frequency interval $[0,1]$ rather than $[-\frac{1}{2},\frac{1}{2}]$. The DFT evaluates the Fourier transform of a signal of length N at N equispaced frequencies $f_k = k/N$, $k=0,1,\ldots,N{-}1$. So we get

$$\widehat{S}_{PER}(f_k) \;=\; \frac{1}{N}\left|\sum_{n=0}^{N-1} y_n\, e^{-j2\pi f_k n}\right|^2 \;=\; \frac{1}{N}\left|\sum_{n=0}^{N-1} y_n\, e^{-j2\pi \frac{k}{N} n}\right|^2 \;,\quad k=0,1,\ldots,N{-}1\;.$$

- We can obtain a finer frequency spacing by padding the data with $N'{-}N$ zeros and then applying an N'-point DFT. The effective data set becomes

$$y'_n \;=\; \begin{cases} y_n \;, & n=0,1,\ldots,N{-}1 \\ 0 \;, & n=N,N{+}1,\ldots,N'{-}1 \end{cases}$$

  which has the same Fourier transform as the original data set. The frequency spacing of the DFT on the data set y'_n will be $\frac{1}{N'} < \frac{1}{N}$. Zero padding: no extra resolution, only a finer evaluation of the periodogram.

The Averaged Periodogram

- The main problem with the periodogram is its large variance. We may introduce an averaging operation in order to reduce the variance.

- Assume we have K independent data records of length L available (K realizations of the same process). So the data in record i are $y_n^{(i)},\; n=0,1,\ldots,L{-}1$ for $i=0,1,\ldots,K{-}1$.

- The *averaged periodogram* is defined as

$$\widehat{S}_{AVPER}(f) \;=\; \frac{1}{K}\sum_{i=0}^{K-1} \widehat{S}^{(i)}_{PER}(f)$$

  where $\widehat{S}^{(i)}_{PER}(f)$ is the periodogram for data record i:

$$\widehat{S}^{(i)}_{PER}(f) \;=\; \frac{1}{L}\left|\sum_{n=0}^{L-1} y_n^{(i)} e^{-j2\pi f n}\right|^2 \;.$$

- Since the data records are i.i.d., the mean of the averaged periodogram is also the mean of the periodogram of any record. Hence

$$E\,\widehat{S}_{AVPER}(f) \;=\; E\,\widehat{S}^{(0)}_{PER}(f) \;=\; W_{B,L}(f) * S_{yy}(f)\;,\quad W_{B,L}(f) = \frac{1}{L}\left(\frac{\sin \pi f L}{\sin \pi f}\right)^2 \;.$$

The Averaged Periodogram (2)

- The variance on the other hand will be decreased by a factor K. Indeed

$$Var\left[\widehat{S}_{AVPER}(f)\right] \;=\; \frac{1}{K} Var\left[\widehat{S}^{(0)}_{PER}(f)\right] \;.$$

- In practice, we normally don't have K independent records but only one record of length N on which to base the spectral estimator. A common approach is to segment the data into K nonoverlapping contiguous blocks of length L so that $N = KL$. In this way, the data records become

$$y^{(i)}_n \;=\; y_{n+iL} \;, \;\; n = 0, 1, \ldots, L{-}1; \;\; i = 0, 1, \ldots, K{-}1 \;.$$

  Contiguous data records cannot be independent unless the y_k are i.i.d. This dependence does not influence the mean of the averaged periodogram. However, the variance reduction obtained by the averaged periodogram is generally less than a factor K. For processes with rapidly decaying acf, the correlation between different data blocks will be weak. If the data are furthermore Gaussian, the data blocks will also be roughly independent.

The Averaged Periodogram (3)

- With the sectioning of a long data record into K shorter data records, the main design issue becomes one of finding a good compromise. Indeed, as the number of data blocks K increases, the variance of the averaged periodogram decreases. However, also the length $L = N/K$ of the data blocks decreases which means that the bandwidth $\approx \frac{1}{L}$ of the Bartlett window increases. This means that the smearing of the spectral estimate (bias) increases.

- One approach is to start with a large value for K (and hence small L) and to observe the averaged periodogram as K decreases. The smearing will decrease and hence more spectral details become apparent. One can stop reducing K when no more details are transpiring. This procedure is called *window closing*. This procedure is not without risk since the variance increases as K decreases and hence spurious details that appear may simply be due to estimation variance.

- We have already discussed before the use of data prewhitening as a technique for reducing the bias problem. This technique may help to alleviate the bias problem in the averaged periodogram.

- If choose e.g. $L = K = \sqrt{N} \;\Rightarrow\;$ AVPER = consistent!

Averaged Periodogram: Welch's Variant

- Welch proposed to
  - introduce a window in the computation of the periodogram corresponding to each data block
  - let the data blocks be overlapping (since the data are windowed, the interaction between consecutive data blocks is less severe than their amount of overlap may suggest). The suggested overlap is as much as 50 or even 75 %.
- Due to the windowing, the spectral leakage gets decreased. Furthermore, due to the overlap, a larger number of data records becomes available, leading to some extra variance reduction.

The Blackman-Tukey Spectral Estimator

- The poor performance of the periodogram may also be illuminated by considering the following equivalent form for the periodogram

$$\begin{aligned}\widehat{S}_{PER}(f) &= \frac{1}{N}\left|\sum_{n=0}^{N-1} y_n\, e^{-j2\pi fn}\right|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} y_n y_m\, e^{-j2\pi f(n-m)} \\ &= \sum_{k=-(N-1)}^{N-1} e^{-j2\pi fk}\,\frac{1}{N}\sum_{n=0}^{N-1-|k|} y_{n+|k|}y_n = \sum_{k=-(N-1)}^{N-1}\hat{r}_{yy}(k)e^{-j2\pi fk}\end{aligned}$$

where

$$\hat{r}_{yy}(k) = \begin{cases}\frac{1}{N}\sum_{n=0}^{N-1-k} y_{n+k}y_n &, \quad k = 0, 1, \ldots, N-1 \\ \hat{r}_{yy}(-k) &, \quad k = -(N-1), \ldots, -1.\end{cases}$$

Hence the periodogram may be seen to be an estimator of the psdf by equivalently estimating the acf and using the Wiener-Khinchin relation.

- The poor performance of the periodogram may be understood by considering the quality of the acf estimate. At lag $N-1$ for instance, we get $\hat{r}_{yy}(N-1) = \frac{1}{N}y_{N-1}y_0$ which apart from being strongly biased is also highly variable due to the lack of averaging. Also, $\hat{r}_{yy}(k) = 0,\ |k| > N-1$.

The Blackman-Tukey Spectral Estimator (2)

- For general lags, the mean of the acf estimate is

$$E\,\hat{r}_{yy}(k) \;=\; \left(1-\frac{|k|}{N}\right) r_{yy}(k) \;=\; w_{B,N}(k) r_{yy}(k)\;,\quad |k| \leq N-1\;.$$

  So the mean is equal to the true value weighted by the Bartlett window and hence the estimator is biased (except for $k=0$).

- We could use an unbiased acf estimator by replacing the $\frac{1}{N}$ factor in $\hat{r}_{yy}(k)$ by $\frac{1}{N-|k|}$. However, this choice leads to a spectral estimate with higher variance. Furthermore, this $\widehat{S}(f)$ may be negative at certain frequencies since the unbiased acf estimate does not necessarily correspond to a positive semidefinite sequence.

- So the estimated acf has a bias that increases linearly with lag and a variance that increases also with lag (number of terms in the averaging decreases). Hence, the trustworthiness of the estimated acf decreases with lag.

The Blackman-Tukey Spectral Estimator (3)

- This motivated Blackman and Tukey to introduce a weighting sequence w_k that reflects the quality of the acf estimates: w_k decreases with $|k|$. This leads to *Blackman-Tukey (BT) spectral estimator*

$$\widehat{S}_{BT}(f) \;=\; \sum_{k=-(N-1)}^{N-1} w_k\,\hat{r}_{yy}(k)\,e^{-j2\pi fk}$$

  where the real sequence w_k (*lag window*) satisfies the following properties:

  1. $0 \leq w_k \leq w_0 = 1$
  2. $w_{-k} = w_k$
  3. $w_k = 0$ for $|k| > M$

  where $M \leq N-1$. Due to this last property, we may rewrite $\widehat{S}_{BT}$ as

$$\widehat{S}_{BT}(f) \;=\; \sum_{k=-M}^{M} w_k\,\hat{r}_{yy}(k)\,e^{-j2\pi fk}$$

- The BT spectral estimator is equivalent to the periodogram if $w_k = 1$ for $|k| \leq M = N-1$. The BT estimator is also sometimes called a *weighted covariance* estimator. The weighting will reduce the variance of the spectral estimate, but again by increasing the bias (smearing).

The Blackman-Tukey Spectral Estimator (4)

- We must be careful however that the window chosen will always lead to a non-negative spectral estimate. Remark that the BT estimator can be rewritten as

  $$\widehat{S}_{BT}(f) \;=\; \mathcal{F}\left\{w_k\,\hat{r}_{yy}(k)\right\} \;=\; W(f) * \widehat{S}_{PER}(f)$$

  since $\mathcal{F}\left\{\hat{r}_{yy}(k)\right\} \;=\; \widehat{S}_{PER}(f)$. Although $\widehat{S}_{PER}(f) \geq 0$, if $W(f)$ is negative at certain frequencies, then the convolution above may produce negative values. To avoid this, it is preferable that $W(f) \geq 0$, $\forall f$ or equivalently that w_k is a nonnegative sequence. Only the Bartlett and Parzen windows (in the table) have nonnegative Fourier transforms.

- We get for the mean of the BT estimator

  $$\begin{aligned} E\,\widehat{S}_{BT}(f) \;&=\; W(f) * E\widehat{S}_{PER}(f) \;=\; W(f) * \left(W_{B,N}(f) * S_{yy}(f)\right) \\ &=\; \underbrace{\left(W(f) * W_{B,N}(f)\right)}_{=\mathcal{F}\left\{w_k\,w_{B,k} \approx w_k\right\}} * S_{yy}(f) \;\approx\; W(f) * S_{yy}(f) \end{aligned}$$

  where we assumed that $N \gg M$. So the true psdf gets smeared again, this time leaving a spectral resolution of about $\frac{1}{M}$. Again, prewhitening the data will help reduce the bias.

The Blackman-Tukey Spectral Estimator (5)

- For the variance, we get under the assumption that the psdf is smooth on a frequency scale of $\frac{1}{M}$, for frequencies not near 0 or $\frac{1}{2}$

  $$Var\left[\widehat{S}_{BT}(f)\right] \;\approx\; S^2_{yy}(f)\,\frac{1}{N}\sum_{k=-M}^{M} w_k^2\;.$$

- Again the bias-variance trade-off becomes apparent: for a small bias, M should be chosen large since that will cause the spectral window $W(f)$ to behave as a Dirac delta function. On the other hand, a small variance imposes a small M. A maximum value of $M = N/5$ is usually recommended. As an example, for the Bartlett window

  $$Var\left[\widehat{S}_{BT}(f)\right] \;\approx\; \frac{2M}{3N}\,S^2_{yy}(f)$$

  so that $M = N/5$ results in a variance reduction by a factor 7.5 compared to the periodogram. Much of the art in classical spectral estimation is in choosing an appropriate window, both the type and the length (M).

- Again, if $M \to \infty$ as $N \to \infty$ s.t. $\frac{N}{M} \to \infty$ (e.g. $M = \sqrt{N}$), then $\widehat{S}_{BT}$ consistent.

The Smoothed Periodogram

- From
$$\widehat{S}_{BT}(f) \;=\; \mathcal{F}\{w_k\,\hat{r}_{yy}(k)\} \;=\; W(f) * \widehat{S}_{PER}(f)$$
we can interpret the BT estimator as the convolution of the periodogram with the spectral window $W(f)$. The role of the spectral window is to smoothen the periodogram, thus possibly increasing the bias but also reducing the variance.

- This suggest computing the periodogram and smoothing it in frequency. Assuming a N' point DFT is used to compute the periodogram, a discrete version with uniform spectral weighting (rectangular $W(f)$) is
$$\widehat{S}_{DAN}(f_k) \;=\; \frac{1}{2L+1}\sum_{i=-L}^{L}\widehat{S}_{PER}(f_k+\frac{i}{N'})\;,\quad f_k=\frac{k}{N'}$$
where we have approximately the correspondence: bandwidth $=\frac{2L+1}{N'}=\frac{1}{M}$. This smoothed periodogram is called *Daniell's spectral estimator*. Many other smoothed periodograms are possible by choosing different spectral weightings. Remark that it is quite easy to control the nonnegativity of the spectral weighting so that a positive psdf estimate can be guaranteed.