# *Part 3. Spectrum Estimation*

# *3.1 Classic Methods for Spectrum Estimation*

*Electrical & Computer Engineering*

*University of Maryland, College Park*

Acknowledgment: ENEE630 slides were based on class notes developed by Profs. K.J. Ray Liu and Min Wu. The slides were made by Prof. Min Wu, with updates from Mr. Wei-Hong Chuang. *Contact: minwu@eng.umd.edu*

# *Logistics*

- Last Lecture: lattice predictor
  - correlation properties of error processes
  - joint process estimator in lattice
  - inverse lattice filter structure
- Today:
  - Spectrum estimation: background and classical methods
- Homework set

# ***Summary of Related Readings on Part-II***

## 2.1 Stochastic Processes and modeling

Haykin (4th Ed) 1.1-1.8, 1.12-1.14

Hayes 3.3 – 3.7 (3.5); 4.7

## 2.2 Wiener filtering

Haykin (4th Ed) Chapter 2

Hayes 7.1, 7.2, 7.3.1

## 2.3-2.4 Linear prediction and Levinson-Durbin recursion

Haykin (4th Ed) 3.1 – 3.3

Hayes 7.2.2; 5.1; 5.2.1 – 5.2.2, 5.2.4– 5.2.5

## 2.5 Lattice predictor

Haykin (4th Ed) 3.8 – 3.10

Hayes 6.2; 7.2.4; 6.4.1

# ***Summary of Related Readings on Part-III***

## Overview    Haykins 1.16, 1.10

## 3.1 Non-parametric method

Hayes 8.1;   8.2 (8.2.3, 8.2.5);   8.3

## 3.2 Parametric method

Hayes 8.5, 4.7;   8.4

## 3.3 Frequency estimation

Hayes 8.6

## Review

– On DSP and Linear algebra: Hayes 2.2, 2.3

– On probability and parameter estimation: Hayes 3.1 – 3.2

# Spectrum Estimation: Background



- Spectral estimation: determine the power distribution in frequency of a random process
  - E.g "Does most of the power of a signal reside at low or high frequencies?" "Are there resonances in the spectrum?"
- Applications:
  - Needs of spectral knowledge in spectrum domain non-causal Wiener filtering, signal detection and tracking, beamforming, etc.
  - Wide use in diverse fields: radar, sonar, speech, biomedicine, geophysics, economics, …
- Estimating p.s.d. of a w.s.s. process is equivalent to estimate autocorrelation at all lags

# ***Spectral Estimation: Challenges***



- When a limited amount of observation data are available
  - Can't get *r(k)* for all k and/or may have inaccurate estimate of *r(k)*
  - Scenario-1: transient measurement (earthquake, volcano, …)
  - Scenario-2: constrained to short period to ensure (approx.) stationarity in speech processing

$$\hat{r}(k) = \frac{1}{N-k} \sum_{n=k+1}^{N} u[n]u^*[n-k], \;\; k = 0,1,\ldots M$$

- Observed data may have been corrupted by noise

# Spectral Estimation:  Major Approaches



- 
  - No assumptions on the underlying model for the data
  - Periodogram and its variations (averaging, smoothing, …)
  - Minimum variance method
- 
  - ARMA, AR, MA models
  - Maximum entropy method
- 
  - For harmonic processes that consist of a sum of sinusoids or complex-exponentials in noise
- 

# Spectral Estimation: Major Approaches

- Nonparametric methods
  - No assumptions on the underlying model for the data
  - Periodogram and its variations (averaging, smoothing, …)
  - Minimum variance method
- Parametric methods
  - ARMA, AR, MA models
  - Maximum entropy method
- Frequency estimation (noise subspace methods)
  - For harmonic processes that consist of a sum of sinusoids or complex-exponentials in noise
- High-order statistics

# Example of Speech Spectrogram



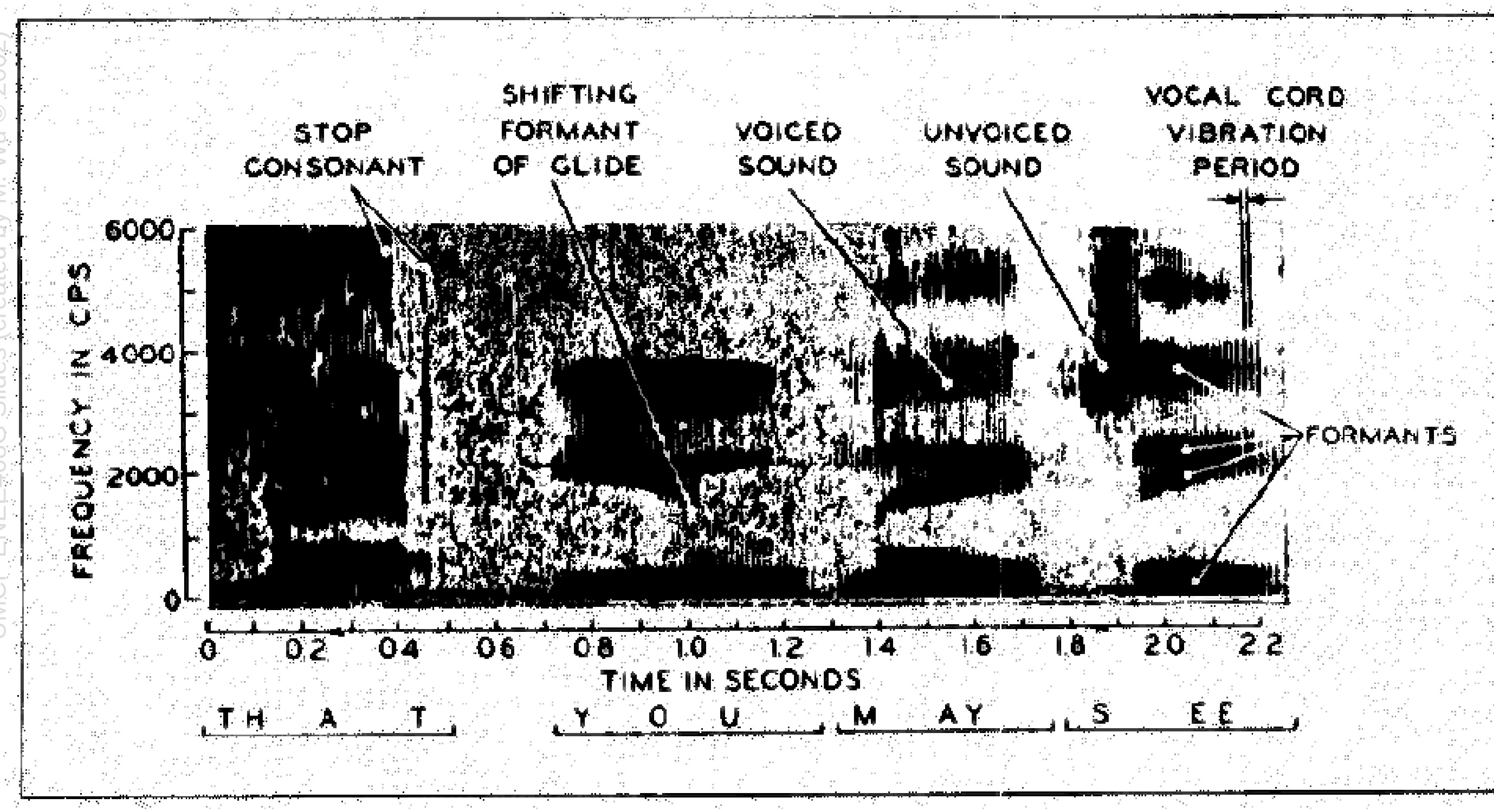
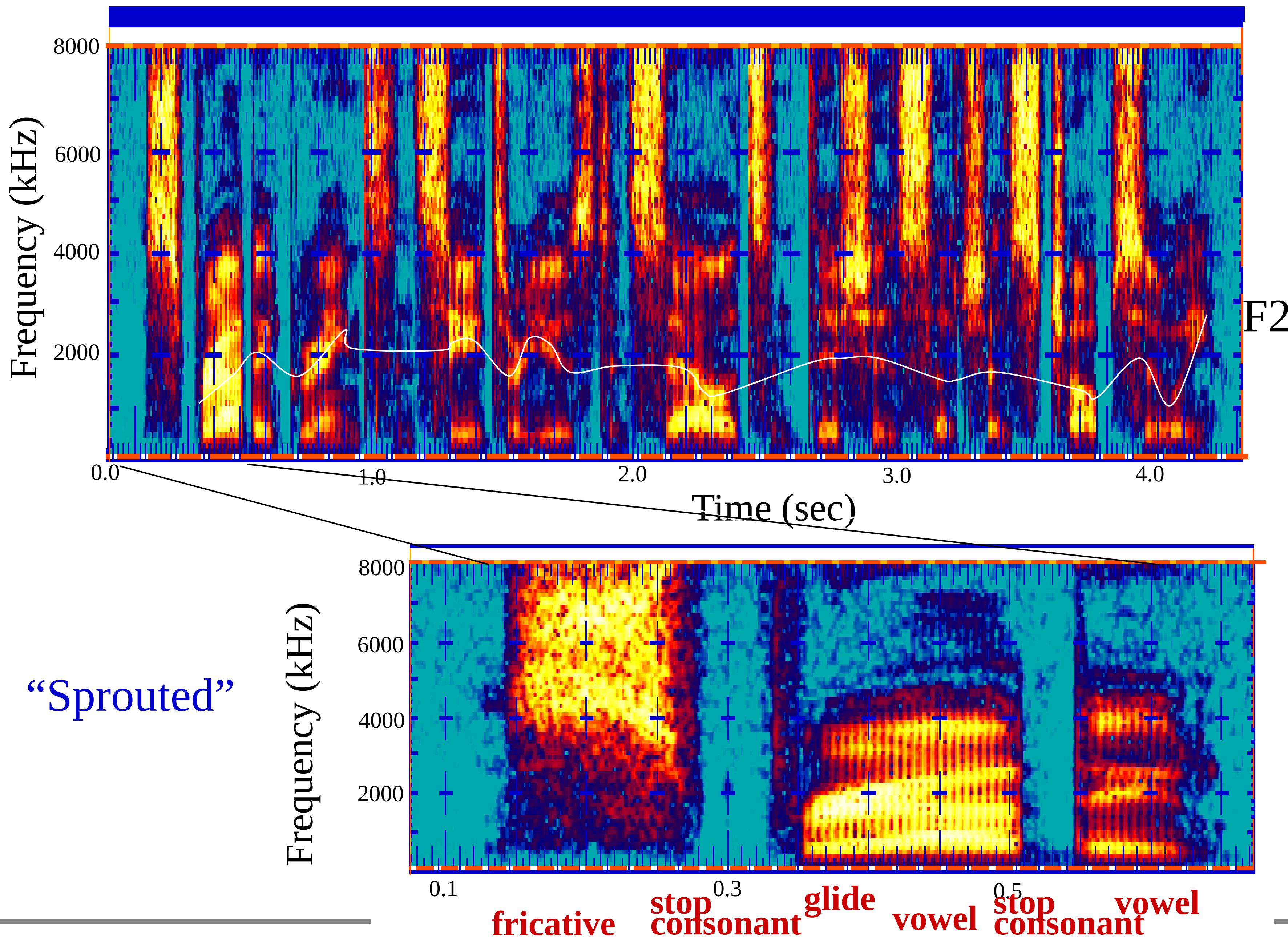



Figure 3 of SPM May'98 Speech Survey

“Sprouted grains and seeds are used in salads and dishes such as chop suey”



# Section 3.1 Classical Nonparametric Methods



Recall: given a w.s.s. process {x[n]} with

$$\begin{cases} E[x[n]] = m_x \\ E[x^*[n]x[n+k]] = r(k) \end{cases}$$

The power spectral density (p.s.d.) is defined as

$$-\frac{1}{2} \le f \le \frac{1}{2}$$

$$(\text{or } \omega = 2\pi f : -\pi \le \omega \le \pi)$$

As we can take DTFT on a specific realization of a random process,

What is the relation between the DTFT of a specific signal and the p.s.d. of the random process?

# Section 3.1 Classical Nonparametric Methods

Recall: given a w.s.s. process {x[n]} with

$$\begin{cases} E[x[n]] = m_x \\ E[x^*[n]x[n+k]] = r(k) \end{cases}$$

The power spectral density (p.s.d.) is defined as

$$P(f) = \sum_{k=-\infty}^{\infty} r(k) e^{-j2\pi fk} \qquad -\frac{1}{2} \le f \le \frac{1}{2}$$

$$(\text{or } \omega = 2\pi f : -\pi \le \omega \le \pi)$$

As we can take DTFT on a specific realization of a random process,
What is the relation between the DTFT of a specific signal and the p.s.d. of the random process?

# Ensemble Average of Squared Fourier Magnitude

- p.s.d. can be related to the ensemble average of the squared Fourier magnitude $|X(\omega)|^2$

$$\text{Consider } P_M(f) \stackrel{\Delta}{=} \frac{1}{2M+1}\left|\sum_{n=-M}^{M} x[n]e^{-j2\pi fn}\right|^2$$

$$= \frac{1}{2M+1}\sum_{n=-M}^{M}\sum_{m=-M}^{M} x[n]x^*[m]e^{-j2\pi f(n-m)}$$

# Ensemble Average of Squared Fourier Magnitude



- p.s.d. can be related to the ensemble average of the squared Fourier magnitude $|X(\omega)|^2$

$$\text{Consider } P_M(f) \overset{\Delta}{=} \frac{1}{2M+1} \left| \sum_{n=-M}^{M} x[n] e^{-j2\pi f n} \right|^2$$

$$= \frac{1}{2M+1} \sum_{n=-M}^{M} \sum_{m=-M}^{M} x[n] x^*[m] e^{-j2\pi f(n-m)}$$

i.e., take DTFT on (2M+1) samples and
examine normalized magnitude

Note: for each frequency f, $P_M(f)$ is a random variable

# Ensemble Average of $P_M(f)$

$$E[P_M(f)] = \frac{1}{2M+1}\sum_{n=-M}^{M}\sum_{m=-M}^{M} r(n-m)e^{-j2\pi f(n-m)}$$

$$= \frac{1}{2M+1}\sum_{k=-2M}^{2M}(2M+1-|k|)r(k)e^{-j2\pi fk}$$

- Now, what if M goes to infinity?

# Ensemble Average of $P_M(f)$

$$E[P_M(f)] = \frac{1}{2M+1} \sum_{n=-M}^{M} \sum_{m=-M}^{M} r(n-m) e^{-j2\pi f(n-m)}$$

$$= \frac{1}{2M+1} \sum_{k=-2M}^{2M} (2M+1-|k|) r(k) e^{-j2\pi fk}$$

$$= \sum_{k=-2M}^{2M} \left(1 - \frac{|k|}{2M+1}\right) r(k) e^{-j2\pi fk}$$

$$= \sum_{k=-2M}^{2M} r(k) e^{-j2\pi fk} - \frac{1}{2M+1} \sum_{k=-2M}^{2M} |k| r(k) e^{-j2\pi fk}$$

- Now, what if M goes to infinity?

# P.S.D. and Ensemble Fourier Magnitude



If the autocorrelation function decays fast enough s.t.

$$\sum_{k=-\infty}^{\infty} |k| r(k) < \infty \quad \text{(i.e., } r(k) \to 0 \text{ rapidly for } k \uparrow \text{)}$$

then $\lim_{M \to \infty} E[P_M(f)] = \sum_{k=-\infty}^{\infty} r(k) e^{-j2\pi fk} = P(f)$

p.s.d.

Thus

# ***P.S.D. and Ensemble Fourier Magnitude***



If the autocorrelation function decays fast enough s.t.

$$\sum_{k=-\infty}^{\infty} |k| r(k) < \infty \quad \text{(i.e., } r(k) \to 0 \text{ rapidly for } k \uparrow)$$

then
$$\lim_{M\to\infty} E[P_M(f)] = \sum_{k=-\infty}^{\infty} r(k) e^{-j2\pi fk} = P(f)$$

p.s.d.

Thus
$$P(f) = \lim_{M\to\infty} E\left[\frac{1}{2M+1}\left|\sum_{n=-M}^{M} x[n] e^{-j2\pi fn}\right|^2\right] \quad (**)$$

## 3.1.1 Periodogram Spectral Estimator



(1) This estimator is based on (**)

Given an observed data set {x[0], x[1], …, x[N-1]}, the periodogram is defined as

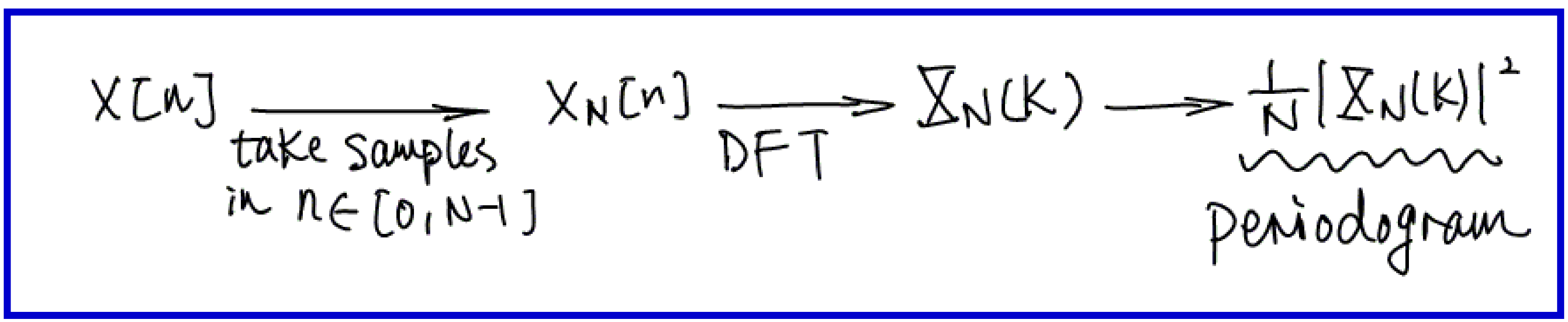
$$\hat{P}_{\mathrm{PER}}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n] e^{-j2\pi fn}\right|^2$$

# 3.1.1 Periodogram Spectral Estimator

(1) This estimator is based on (**)

Given an observed data set {x[0], x[1], …, x[N-1]}, the periodogram is defined as

$$\hat{P}_{\mathrm{PER}}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n] e^{-j2\pi f n}\right|^2$$



# An Equivalent Expression of Periodogram

The periodogram estimator can be given in terms of $\hat{r}(k)$

$$\hat{P}_{\mathrm{PER}}(f) = \sum_{k=-(N-1)}^{N-1} \hat{r}(k) e^{-j2\pi f k}$$

where $\hat{r}(k) =$

– The quality of the estimates for the higher lags of r(k) may be poorer since they involve fewer terms of lag products in the averaging operation

Exercise: to show this from the periodogram definition in last page

# An Equivalent Expression of Periodogram

The periodogram estimator can be given in terms of $\hat{r}(k)$

$$\hat{P}_{\mathrm{PER}}(f) = \sum_{k=-(N-1)}^{N-1} \hat{r}(k) e^{-j2\pi f k}$$

where $\hat{r}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x^*[n] x[n+k]; \; \hat{r}(-k) = \hat{r}^*(k) \text{ for } k \geq 0$

– The quality of the estimates for the higher lags of r(k) may be poorer since they involve fewer terms of lag products in the averaging operation

Exercise: to show this from the periodogram definition in last page

# (2) Filter Bank Interpretation of Periodogram

For a particular frequency of $f_0$:

$$\hat{P}_{\mathrm{PER}}(f_0) = \frac{1}{N}\left|\sum_{k=0}^{N-1} e^{-j2\pi f_0 k} x[k]\right|^2$$

$$= \left[ N \cdot \left|\sum_{k=0}^{N-1} h[n-k]x[k]\right|^2 \right]_{n=0}$$

where

$$h[n] =$$

– Impulse response of the filter h[n]: a windowed version of a complex exponential

# (2) Filter Bank Interpretation of Periodogram



For a particular frequency of $f_0$:

$$\hat{P}_{\mathrm{PER}}(f_0) = \frac{1}{N}\left|\sum_{k=0}^{N-1} e^{-j2\pi f_0 k} x[k]\right|^2$$

$$= \left[ N \cdot \left|\sum_{k=0}^{N-1} h[n-k]x[k]\right|^2 \right]_{n=0}$$

where

$$h[n] = \begin{cases} \frac{1}{N}\exp(j2\pi f_0 n) & \text{for } n = -(N-1),\ldots,-1,0; \\ 0 & \text{otherwise} \end{cases}$$

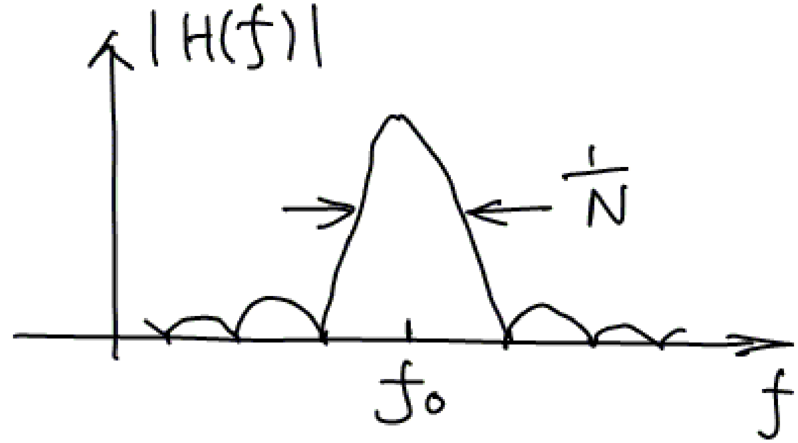
– Impulse response of the filter h[n]: a windowed version of a complex exponential

# Frequency Response of h[n]



$$H(f) = \underbrace{\frac{\sin N\pi(f - f_0)}{N \sin \pi(f - f_0)}} \exp\left[j(N-1)\pi(f - f_0)\right]$$

sinc-like function centered at $f_{0:}$



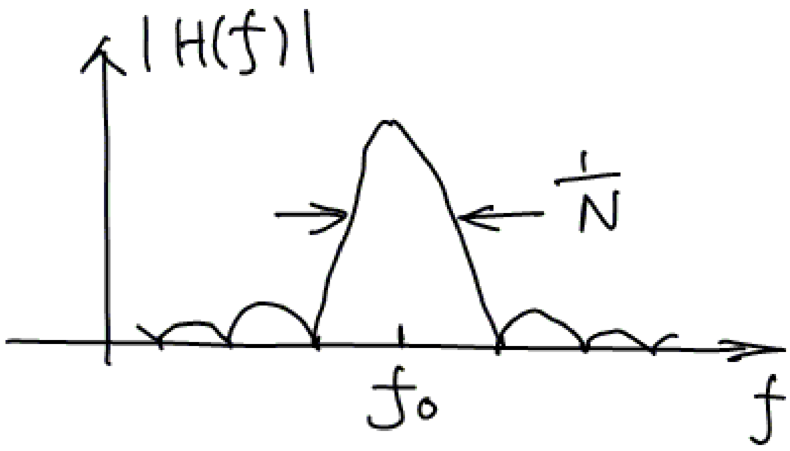
# Frequency Response of h[n]



$$H(f) = \underbrace{\frac{\sin N\pi(f - f_0)}{N \sin \pi(f - f_0)}} \exp\left[j(N-1)\pi(f - f_0)\right]$$

sinc-like function centered at $f_{0:}$

- H(f) is a bandpass filter
  - Center frequency is $f_0$
  - 3dB bandwidth ≈ 1/N



# ***Periodogram: Filter Bank Perspective***



- Can view the periodogram as an estimator of power spetrum that has a built-in filterbank
  - The filter bank ~ a set of bandpass filters

$$\hat{P}_{\mathrm{PER}}(f_0) = \left[ N \cdot \left| \sum_{k=0}^{N-1} h[n-k]x[k] \right|^2 \right]_{n=0}$$

# Periodogram: Filter Bank Perspective

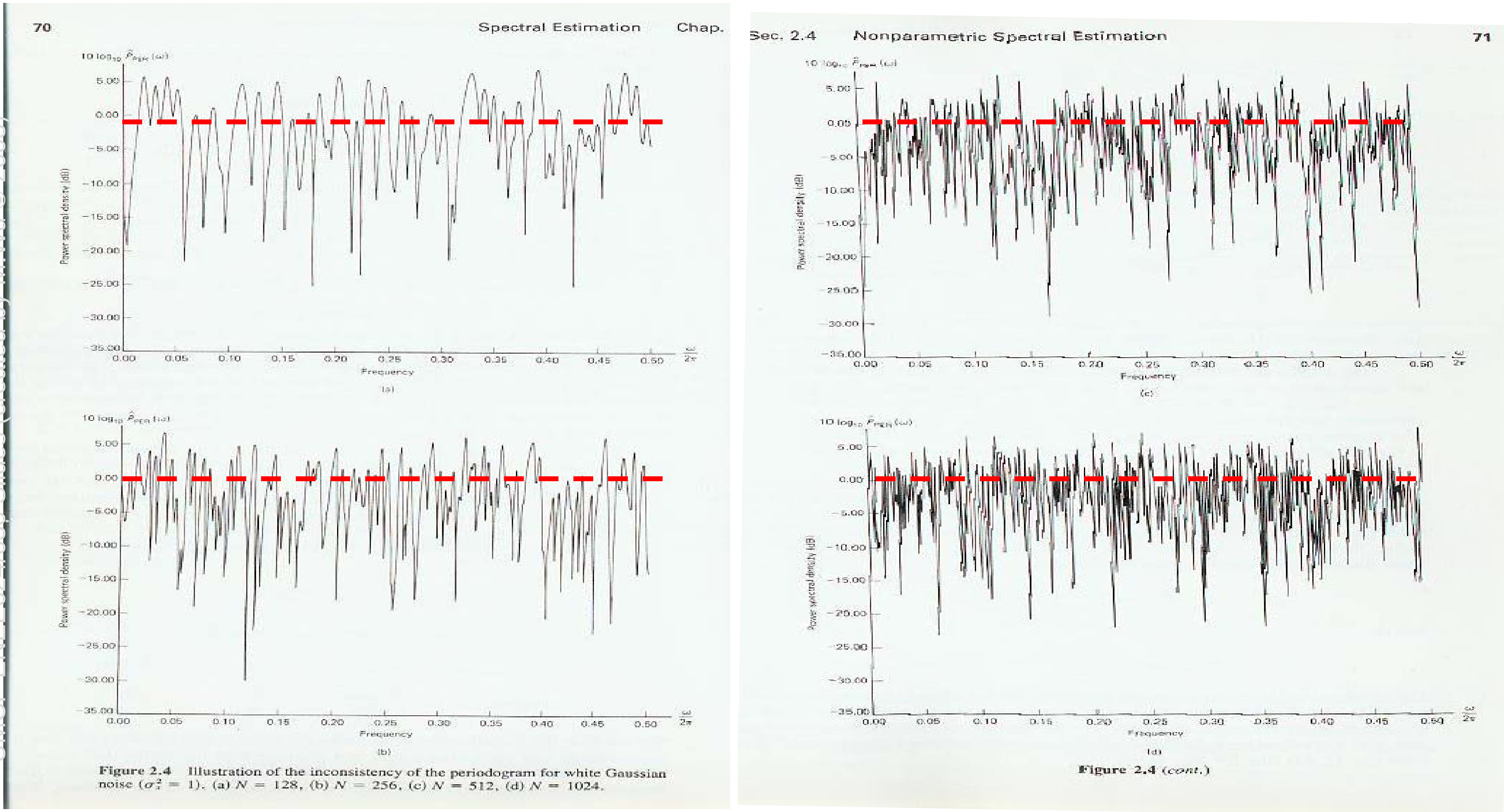


- Can view the periodogram as an estimator of power spetrum that has a built-in filterbank
  - The filter bank ~ a set of bandpass filters
  - The estimated p.s.d. for each frequency $f_0$ is the power of one output sample of the bandpass filter centering at $f_0$

$$\hat{P}_{\text{PER}}(f_0) = \left[ N \cdot \left| \sum_{k=0}^{N-1} h[n-k]x[k] \right|^2 \right]_{n=0}$$

# E.g. White Gaussian Process

[Lim/Oppenheim Fig.2.4]
Periodogram of zero-mean white Gaussian noise using N-point data record: N=128, 256, 512, 1024




- The random fluctuation (measured by variance) of the periodogram does not decrease with increasing N
  ➔ periodogram is not a consistent estimator

# (3) How Good is Periodogram for Spectral Estimation?

$$\text{If } N \to \infty, \text{ will } \hat{P}_{\text{PER}} \to \text{p.s.d. } P(f)?$$

- Estimation: Tradeoff between bias and variance

$$E(\hat{\theta}) \overset{?}{=} \theta$$

$$E[\,|\hat{\theta} - E(\hat{\theta})|^2\,] = ?$$

- For white Gaussian process, we can show that at $f_k = k/N$

$$\Rightarrow \quad E[\hat{P}_{PER}(f_k)] = P(f_k), \quad k = 0, 1, \ldots N/2$$

$$Var[\hat{P}_{PER}(f_k)] = \begin{cases} P^2(f_k), & k = 1, \ldots \frac{N}{2}-1 \\ 2P^2(f_k), & k = 0, \frac{N}{2} \end{cases} \quad \propto P^2(f_k)$$

# ***Performance of Periodogram: Summary***



- The periodogram for white Gaussian process is an unbiased estimator but not consistent
  - The variance does not decrease with increasing data length
  - Its standard deviation is as large as the mean (equal to the quantity to be estimated)
- Reasons for the poor estimation performance
  - Given N real data points, the # of unknown parameters $\{P(f_0), \ldots P(f_{N/2})\}$ we try to estimate is N/2, i.e. proportional to N
- Similar conclusions can be drawn for processes with arbitrary p.s.d. and arbitrary frequencies
  - Asymptotically unbiased (as N goes to infinity) but inconsistent

# 3.1.2 Averaged Periodogram



- As one solution to the variance problem of periodogram
  - Average K periodograms computed from K sets of data records

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

where $$\hat{P}^{(m)}_{\text{PER}}(f) = \frac{1}{L}\left|\sum_{n=0}^{L-1} x_m[n] e^{-j2\pi fn}\right|^2$$

And the K sets of data records are

$$\{x_0[0], \ldots, x_0[L-1];\ x_1[n],\ 0 \le n \le L-1; \ldots$$
$$\{x_{K-1}[n-1],\ 0 \le n \le L-1\}$$

# Performance of Averaged Periodogram

– If K sets of data records are uncorrelated with each other, we have: $(f_i = i/L)$

$\hat{P}_{PER}^{(m)}(f)$ i.i.d. (m=0,1, … L-1) for white Gaussian process

$$\Rightarrow Var[\hat{P}_{AVPER}(f)] = \qquad \propto \frac{1}{K} P^2(f_i)$$

# Performance of Averaged Periodogram

– If K sets of data records are uncorrelated with each other, we have : $(f_i = i/L)$

$\hat{P}_{PER}^{(M)}(f)$ i.i.d. (m=0,1, … L-1) for white Gaussian process

$$\Rightarrow \mathrm{Var}[\hat{P}_{AVPER}(f)] = \begin{cases} \frac{1}{K} P^2(f_i) & i = 1, 2, \cdots \frac{L}{2}-1 \\ \frac{2}{K} P^2(f_i) & i = 0, \frac{L}{2} \end{cases}$$

$$\propto \frac{1}{K} P^2(f_i)$$

i.e., $K \uparrow \to \mathrm{Var} \downarrow$, and $\mathrm{Var} \to 0$ for $K \to \infty$

i.e., consistent estimate

# Practical Averaged Periodogram

- Usually we partition an available data sequence of length N
  - into K non-overlapping blocks, each block has length L (i.e. N=KL)

i.e. $$x_m[n] = x[n+mL], \qquad n = 0, 1, \dots, L-1$$
$$m = 0, 1, \dots, K-1$$

- Since the blocks are contiguous, the K sets of data records may not be completely uncorrelated
  - Thus the variance reduction factor is in general less than K

- Periodogram averaging is also known as the Bartlett's method

# *Averaged Periodogram for Fixed Data Size*

- Given a data record of fixed size N, will the result be better if we segment the data into more and more subrecords?

We examine for a real-valued stationary process:

$$E\left[\hat{P}_{\text{AV PER}}(f)\right] = E\left[\frac{1}{K}\sum_{m=0}^{K-1}\hat{P}^{(m)}_{\text{PER}}(f)\right] = E\left[\hat{P}^{(0)}_{\text{PER}}(f)\right]$$

identical distribution for all *m*

Note
$$\hat{P}^{(0)}_{\text{PER}}(f) = \sum_{l=-(L-1)}^{L-1}\hat{r}^{(0)}(l)e^{-j2\pi fl}$$

where
$$\hat{r}^{(0)}(l) = \frac{1}{L}\sum_{n=0}^{L-1-|l|}x[n]x\left[n+|l|\right]$$

$\rightarrow$ an equivalent expression to definition in terms of x[n]

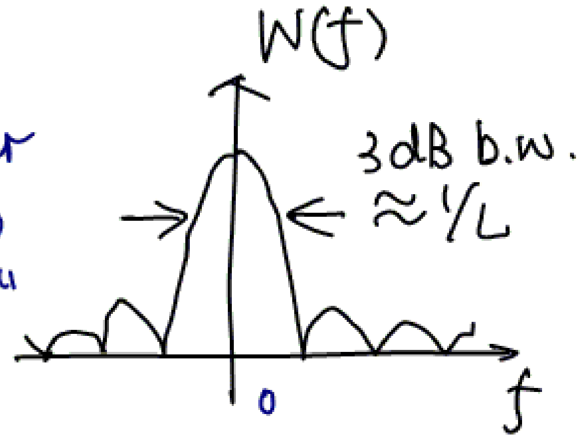
# ***Mean of Averaged Periodogram***



$$W[k] = \begin{cases} 1 - |k|/L & \text{for } |k| \le L-1 \\ 0 & \text{o.w.} \end{cases}$$

"triangular (Bartlett) window"

$$\Rightarrow W(f) = \frac{1}{L}\left(\frac{\sin \pi f L}{\sin \pi f}\right)^2$$



# Mean of Averaged Periodogram



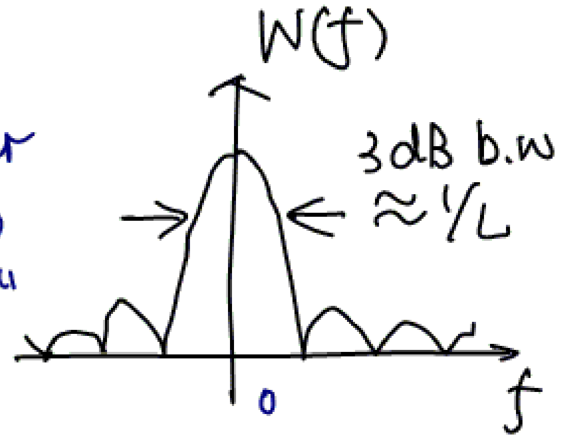
$$\Rightarrow E[\hat{r}^{(0)}(\ell)] = \underbrace{\left(1 - \frac{|\ell|}{L}\right)}_{\triangleq\, w[\ell]} r(\ell) \quad \text{for } |\ell| \le L-1$$

$$\therefore E[\hat{P}_{AVPER}(f)] = \sum_{\ell=-(L-1)}^{L-1} w[\ell]\, r(\ell)\, e^{-j2\pi f \ell}$$

$$w[k] = \begin{cases} 1 - |k|/L & \text{for } |k| \le L-1 \\ 0 & \text{o.w.} \end{cases}$$

"triangular (Bartlett) window"

$$\Rightarrow W(f) = \frac{1}{L}\left(\frac{\sin \pi f L}{\sin \pi f}\right)^2$$



# Mean of Averaged Periodogram (cont'd)

$$E[\hat{P}_{\text{AV PER}}(f)] = \text{DTFT}\left[\{w[k]r(k)\}\right]\Big|_f$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} W(f-\eta)P(\eta)\,d\eta$$

$$\neq P(f)$$

multiplication in time

↓

convolution in frequency

- Biased estimate (both averaged and regular periodogram)
  - The convolution with the window function w[k] lead to the mean of the averaged periodogram being smeared from the true p.s.d

# Mean of Averaged Periodogram (cont'd)

$$E[\hat{P}_{\text{AV PER}}(f)] = \text{DTFT}\left[\{w[k]r(k)\}\right]\Big|_f$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} W(f-\eta)P(\eta)d\eta$$

$$\neq P(f)$$

multiplication in time

↓

convolution in frequency

- Biased estimate (both averaged and regular periodogram)
  - The convolution with the window function w[k] lead to the mean of the averaged periodogram being smeared from the true p.s.d
- Asymptotic unbiased as L $\rightarrow \infty$
  - To avoid the smearing, the window length L must be large enough so that the narrowest peak in P(f) can be resolved
- This gives a tradeoff between bias and variance

Small K => better resolution (smaller smearing/bias) but larger variance

# ***Non-parametric Spectrum Estimation: Recap***

- Periodogram
  - Motivated by relation between p.s.d. and squared magnitude of DTFT of a finite-size data record
  - Variance: won't vanish as data length N goes infinity ~ "inconsistent"
  - Mean: asymptotically unbiased w.r.t. data length N in general
    - *equivalent to apply triangular window to autocorrelation function (windowing in time gives smearing/smoothing in freq domain)*
    - *unbiased for white Gaussian*
- Averaged periodogram
  - Reduce variance by averaging K sets of data record of length L each
  - Small L increases smearing/smoothing in p.s.d. estimate thus higher bias ➔ *equiv. to triangular windowing*
- Windowed periodogram: generalize to other symmetric windows

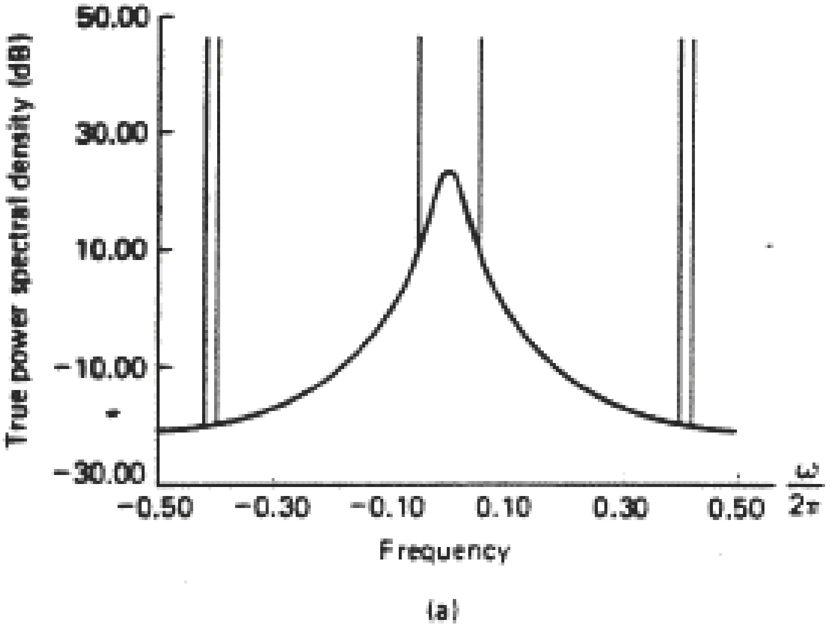
# ***Case Study on Non-parametric Methods***

- Test case: a process consists of narrowband components (sinusoids) and a broadband component (AR)
  - $x[n] = 2\cos(\omega_1 n) + 2\cos(\omega_2 n) + 2\cos(\omega_3 n) + z[n]$

    where $z[n] = -a_1 z[n-1] + v[n]$, $a_1 = -0.85$, $\sigma^2 = 0.1$

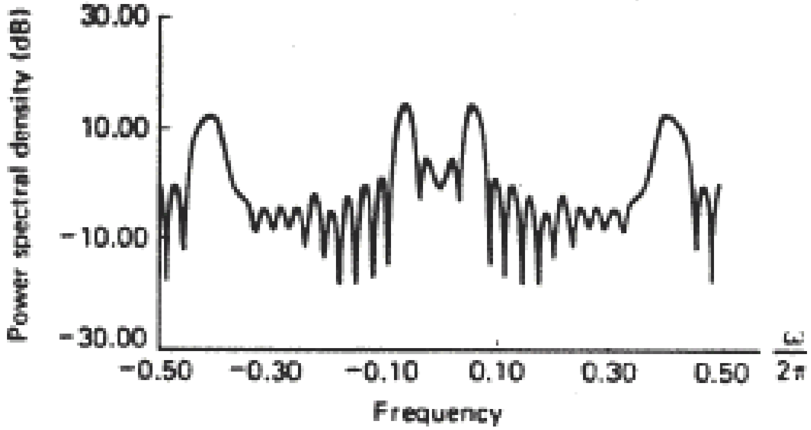
    $$\omega_1/2\pi = 0.05,\ \omega_2/2\pi = 0.40,\ \omega_3/2\pi = 0.42$$
  - N=32 data points are available

    ➔ periodogram resolution f = 1/32



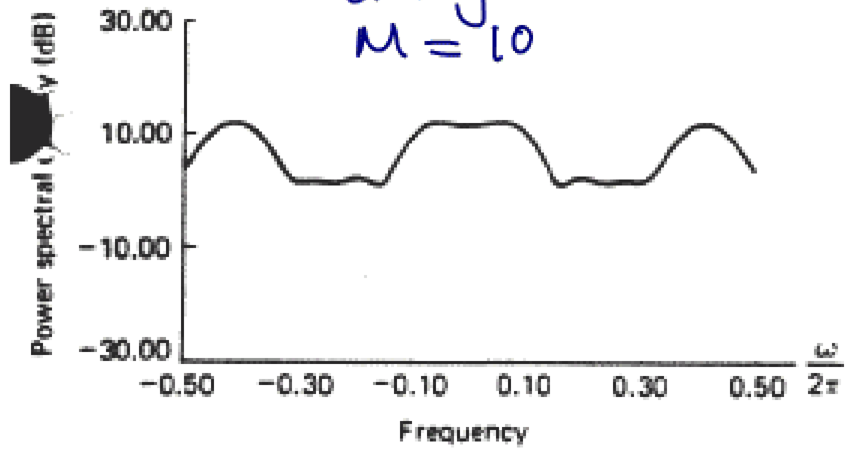

- Examine typical characteristics of various non-parametric spectral estimators

(Fig.2.17 from Lim/Oppenheim book)

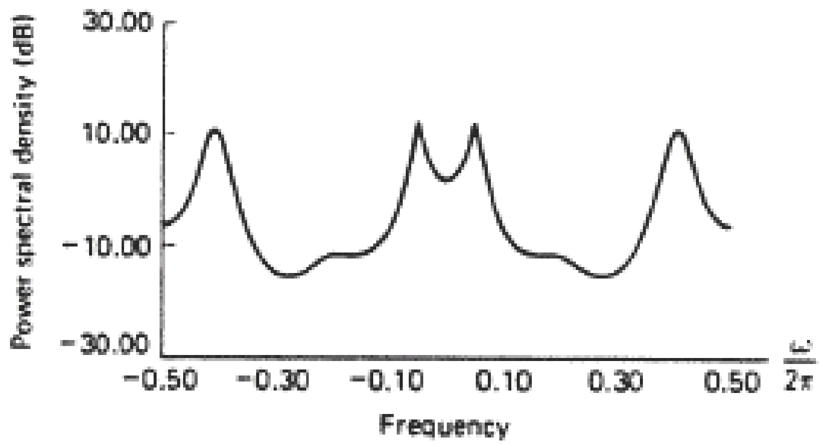



(b) Periodogram

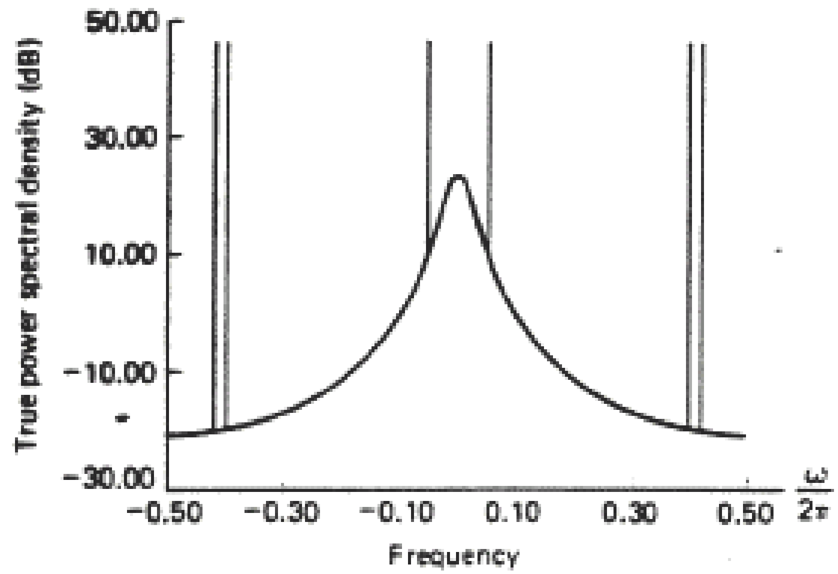
triangle window
M = 10




(c) Blackman-Tukey




(d) Minimum variance spectral estimator




true p.s.d.

(a)

# 3.1.3 Periodogram with Windowing

- Review and Motivation

The periodogram estimator can be given in terms of $\hat{r}(k)$

$$\hat{P}_{\text{PER}}(f) = \sum_{k=-(N-1)}^{N-1} \hat{r}(k) e^{-j2\pi fk}$$

where $$\hat{r}(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x^*[n] x[n+k]; \quad \hat{r}(-k) = \hat{r}^*(k) \quad \text{for } k \geq 0$$

  - The higher lags of r(k), the poorer estimates since the estimates involve fewer terms of lag products in the averaging operation

- Solution: weigh the higher lags less
  - Trade variance with bias

# ***Windowing***

- Use a window function to weigh the higher lags less

i.e. $$\hat{P}_{win}(f) = \sum_{k=-(N-1)}^{N-1} w[k]\, \hat{r}(k)\, e^{-j2\pi f k}$$

where $w[k]$ is a "lag window" with properties of:

① $0 \le w[k] \le w[0] = 1$

w(0)=1 preserves variance r(0)

② $w[-k] = w[k]$ symmetric

③ $w[k] = 0$ for $|k| > M$ where $M \le N-1$

④ $W(f)$ must be chosen to ensure $\hat{P}_{win}(f) \ge 0$

- Effect: periodogram smoothing
  - Windowing in time ⇔ Convolution/filtering the periodogram
  - Also known as the Blackman-Tukey method

# Common Lag Windows

- Much of the art in non-parametric spectral estimation is in choosing an appropriate window (both in type and length)

TABLE 2.1 COMMON LAG WINDOWS

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

Table 2.1 common lag window (from Lim-Oppenheim book)

# ***Discussion: Estimate r(k) via Time Average***

- Normalizing the sum of (N-k) pairs

by a factor of 1/N ? v.s. by a factor of 1/(N-k) ?

<u>Biased</u> (low variance) <u>Unbiased</u> (may not non-neg. definite)

$$\hat{r}_1(k) = \frac{1}{N}\sum_{n=0}^{N-1-k} x[n+k]\,x^*[n];\qquad \hat{r}_2(k) = \frac{1}{N-k}\sum_{n=0}^{N-1-k} x[n+k]\,x^*[n]$$

$$E(\hat{r}_1(k)) = \qquad\qquad E(\hat{r}_2(k)) =$$

- Hints on showing the non-negative definiteness: using $\hat{r}_1(k)$ to construct correlation matrix
- For $\hat{r}_2(k)$ : HW#8

# Discussion: Estimate r(k) via Time Average

- Normalizing the sum of (N-k) pairs

by a factor of 1/N ? v.s. by a factor of 1/(N-k) ?

<u>Biased</u> (low variance) <u>Unbiased</u> (may not non-neg. definite)

$$\hat{r}_1(k) = \frac{1}{N}\sum_{n=0}^{N-1-k} x[n+k]\,x^*[n]; \quad \hat{r}_2(k) = \frac{1}{N-k}\sum_{n=0}^{N-1-k} x[n+k]\,x^*[n]$$

$$E(\hat{r}_1(k)) = \frac{N-k}{N} r(k) \qquad E(\hat{r}_2(k)) = r(k)$$

- Hints on showing the non-negative definiteness: using $\hat{r}_1(k)$ to construct correlation matrix

$\hat{R}_N = \mathcal{X}^H \mathcal{X}$, where

$$\mathcal{X} = \frac{1}{\sqrt{N}}\begin{bmatrix} x(0) & 0 & 0 & & \\ x(1) & x(0) & 0 & & \\ \vdots & \vdots & x(0) & \ddots & x(0) \\ x(N-1) & \vdots & \vdots & & \vdots \\ 0 & x(N-1) & \vdots & & \vdots \\ \vdots & 0 & & & \vdots \\ & \vdots & & & \end{bmatrix}$$

- For $\hat{r}_2(k)$ : HW#8

## *3.1.4 Minimum Variance Spectral Estimation (MVSE)*

- Recall: filter bank perspective of periodogram
  - The periodogram can be viewed as estimating the p.s.d. by forming a bank of narrowband filters with sinc-like response
  - The high sidelobe can lead to "leakage" problem:
    - *large output power due to p.s.d outside the band of interest*

- MVSE designs filters to minimize the leakage from out-of-band spectral components
  - Thus the shape of filter is dependent on the frequency of interest and data adaptive

    (unlike the identical filter shape for periodogram)

  - MVSE is also referred to as the Capon spectral estimator

# ***Main Steps of MVSE Method***

- Design a bank of bandpass filters $H_i(f)$ with center frequency $f_i$ so that
  - Each filter rejects the maximum amount of out-of-band power
  - And passes the component at frequency $f_i$ without distortion

- Filter the input process { x[n] } with each filter in the filter bank and estimate the power of each output process

- Set the power spectrum estimate at frequency $f_i$ to be the power estimated above divided by the filter bandwidth

## ***Formulation of MVSE***

The MVSE designs a filter H(f) for each frequency of interest $f_0$

minimize the output power

(i.e., to pass the components at $f_0$ w/o distortion)

# ***Formulation of MVSE***

The MVSE designs a filter H(f) for each frequency of interest $f_0$

minimize the output power

$$\rho = \int_{-\frac{1}{2}}^{+\frac{1}{2}} |H(f)|^2 P(f) df$$

subject to $$H(f_0) = 1$$

(i.e., to pass the components at $f_0$ w/o distortion)

# ***Deriving MVSE Solutions***

# Output Power From H(f) filter

From the filter bank perspective of periodogram:

$$H(f) = \sum_{n=-(N-1)}^{0} h[n] e^{-j2\pi f n}$$

Thus

$$\rho = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{k=-(N-1)}^{0} h[k] e^{-j2\pi f k} \sum_{l=-(N-1)}^{0} h^*[l] e^{j2\pi f l} P(f) df$$

Equiv. to filter r(k) with { h(k) ⊗ h*(-k) } and evaluate at output time k=0

# Output Power From H(f) filter

From the filter bank perspective of periodogram:

$$H(f) = \sum_{n=-(N-1)}^{0} h[n] e^{-j2\pi fn}$$

Thus

$$\rho = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{k=-(N-1)}^{0} h[k] e^{-j2\pi fk} \sum_{l=-(N-1)}^{0} h^*[l] e^{j2\pi fl} P(f) df$$

$$= \sum_{k=-(N-1)}^{0} \sum_{l=-(N-1)}^{0} h[k] h^*[l] \int_{-\frac{1}{2}}^{\frac{1}{2}} P(f) e^{j2\pi f(l-k)} df$$

$$= \sum_{k=-(N-1)}^{0} \sum_{l=-(N-1)}^{0} h[k] h^*[l] r(l-k)$$

Equiv. to filter r(k) with { h(k) ⊗ h*(-k) } and evaluate at output time k=0

# *Matrix-Vector Form of MVSE Formulation*

Define

$$\underline{h}^* \triangleq \begin{bmatrix} h[0] \\ h[-1] \\ \vdots \\ h[-(N-1)] \end{bmatrix} \Rightarrow P = \underline{h}^H R^T \underline{h}$$

$$[h(0), h(-1) \ldots h(1-N)] \begin{bmatrix} r(0) & r(-1) & \cdots \\ r(1) & r(0) & \\ \vdots & & \ddots \end{bmatrix} \begin{bmatrix} h(0) \\ \vdots \\ \vdots \end{bmatrix}^*$$

$$\underline{e} = \begin{bmatrix} 1 \\ e^{j2\pi f_0} \\ \vdots \\ e^{j2\pi(N-1)f_0} \end{bmatrix}$$

➔ The constraint can be written in vector form as $\underbrace{\underline{h}^H \underline{e}}_{H(f_0)} = 1$

# Matrix-Vector Form of MVSE Formulation

Define

$$\underline{h}^* \triangleq \begin{bmatrix} h[0] \\ h[-1] \\ \vdots \\ h[-(N-1)] \end{bmatrix} \Rightarrow P = \underline{h}^H R^T \underline{h}$$

$$[h(0), h(-1) \ldots h(1-N)] \begin{bmatrix} r(0) & r(-1) & \cdots \\ r(1) & r(0) & \\ \vdots & & \ddots \end{bmatrix} \begin{bmatrix} h^*(0) \\ \vdots \\ \vdots \end{bmatrix}$$

$$\underline{e} = \begin{bmatrix} 1 \\ e^{j2\pi f_0} \\ \vdots \\ e^{j2\pi(N-1)f_0} \end{bmatrix}$$

➔ The constraint can be written in vector form as $\underbrace{\underline{h}^H \underline{e}}_{H(f_0)} = 1$

Thus the problem becomes

$$\min_{\underline{h}} \underline{h}^H R^T \underline{h} \qquad \text{subject to} \quad \underline{h}^H \underline{e} = 1$$

## Solution of MVSE

$$J \overset{def}{=} \underline{h}^H R^T \underline{h} + \mathrm{Re}\left[2\lambda(1-\underline{h}^H \underline{e})\right]$$

- Use Lagrange multiplier approach for solving the constrained optimization problem
  - Define **real-valued** objective function s.t. the stationary condition can be derived in a simple and elegant way based on the theorem for complex derivative/gradient operators

$$\min_{\underline{h},\lambda} J = \underline{h}^H R^T \underline{h} + \lambda(1-\underline{h}^H \underline{e}) + \left[\lambda(1-\underline{h}^H \underline{e})\right]^*$$

$$= \underline{h}^H R^T \underline{h} + \lambda(1-\underline{h}^H \underline{e}) + \lambda^*(1-\underline{e}^H \underline{h})$$

either $\nabla_{\underline{h}^*} J = 0 \Rightarrow R^T \underline{h} - \lambda \underline{e} = 0$

or $\nabla_{\underline{h}} J = 0 \Rightarrow \left(\underline{h}^H R^T\right)^T - \lambda^* \underline{e}^* = 0$

$$\Rightarrow \left(R^T\right)^H \underline{h} - \lambda \underline{e} = 0 \Rightarrow R^T \underline{h} - \lambda \underline{e} = 0$$

$$\Rightarrow \underline{h} = \lambda\left(R^T\right)^{-1} \underline{e} \text{ and } \underline{h}^H \underline{e} = 1$$

$$\Rightarrow \begin{cases} \lambda = \dfrac{1}{\underline{e}^H \left(R^T\right)^{-1} \underline{e}} \\ h = \dfrac{\left(R^T\right)^{-1} \underline{e}}{\underline{e}^H \left(R^T\right)^{-1} \underline{e}} \end{cases}$$

# Solution of MVSE (cont'd)

The optimal filter:

$$\underline{h} = \frac{\left(R^T\right)^{-1}\underline{e}}{\underline{e}^H\left(R^T\right)^{-1}\underline{e}}$$

It follows that

$$\rho = \underline{h}^H R^T \underline{h} = \underline{h}^H \lambda R^T \left(R^T\right)^{-1}\underline{e}$$

$$= \lambda \underline{h}^H \underline{e} = \lambda = \frac{1}{\underline{e}^H\left(R^T\right)^{-1}\underline{e}}$$

# MVSE: Summary

If choosing the bandpass filters to be FIR of length p, its 3dB-b.w. is approximately 1/p

Thus the MVSE is

$$\hat{P}_{\mathrm{MV}}(f) = \frac{p}{\underline{e}^H \left(\hat{R}^T\right)^{-1} \underline{e}}$$

(i.e. normalize by filter b.w.)

$\hat{R}$ is $p \times p$ correlation matrix

$$\underline{e} = \begin{bmatrix} 1 \\ \exp(j2\pi f) \\ \vdots \\ \exp(j2\pi f(p-1)) \end{bmatrix}$$

- MVSE is a data adaptive estimator and provides improved resolution over periodogram
  - Also referred to as “High-Resolution Spectral Estimator”
  - Does not assume a particular underlying model for the data

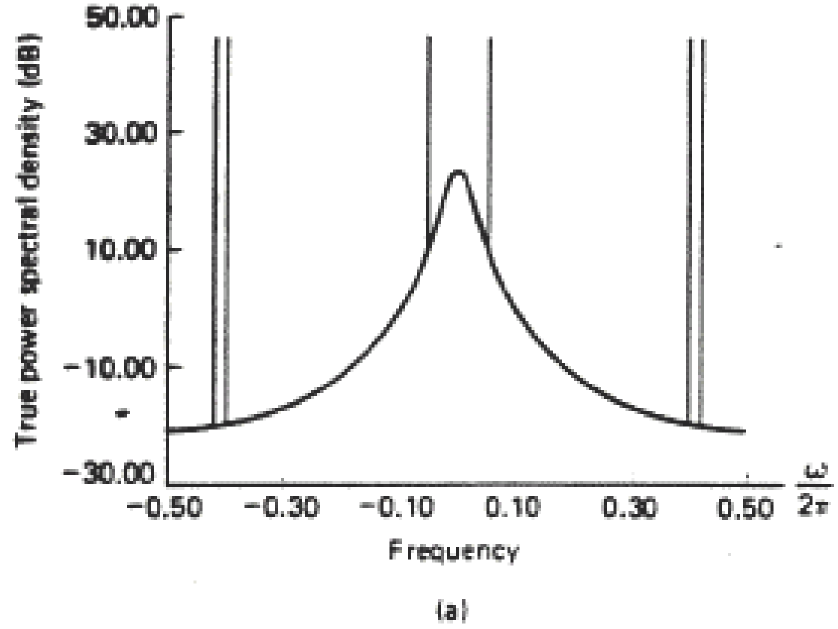
# Recall: Case Study on Non-parametric Methods

- Test case: a process consists of narrowband components (sinusoids) and a broadband component (AR)
  - $x[n] = 2\cos(\omega_1 n) + 2\cos(\omega_2 n) + 2\cos(\omega_3 n) + z[n]$

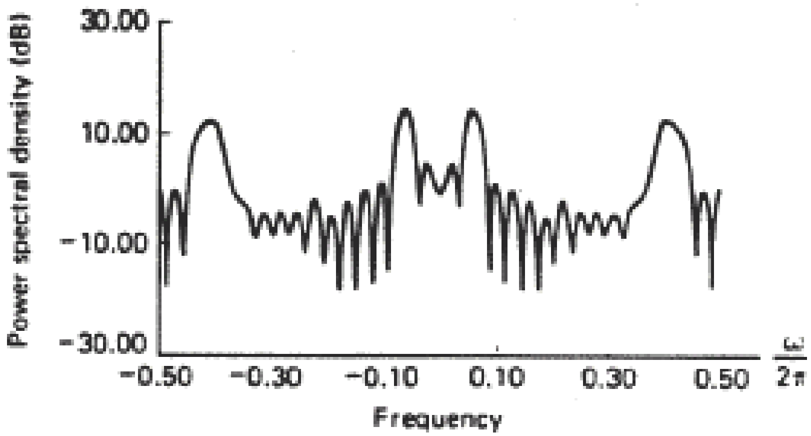
    where $z[n] = -a_1 z[n-1] + v[n]$, $a_1 = -0.85$, $\sigma^2 = 0.1$

    $\omega_1/2\pi = 0.05$, $\omega_2/2\pi = 0.40$, $\omega_3/2\pi = 0.42$
  - N=32 data points are available
    ➔ periodogram resolution f = 1/32

- Examine typical characteristics of various non-parametric spectral estimators

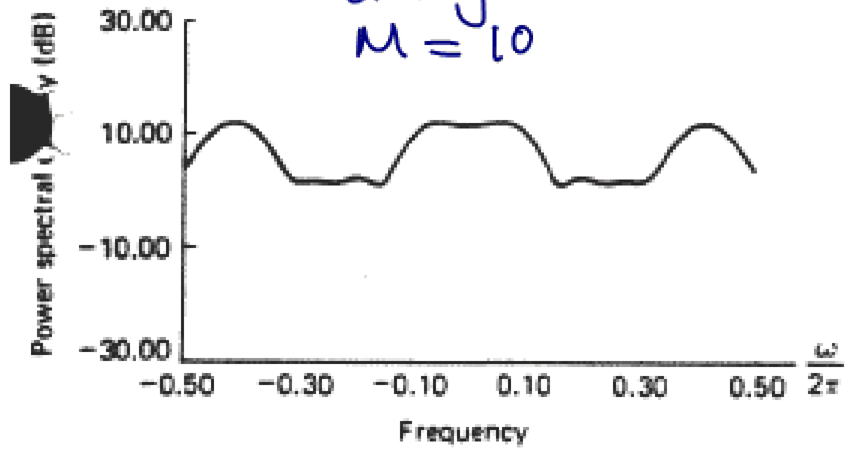



(Fig.2.17 from Lim/Oppenheim book)

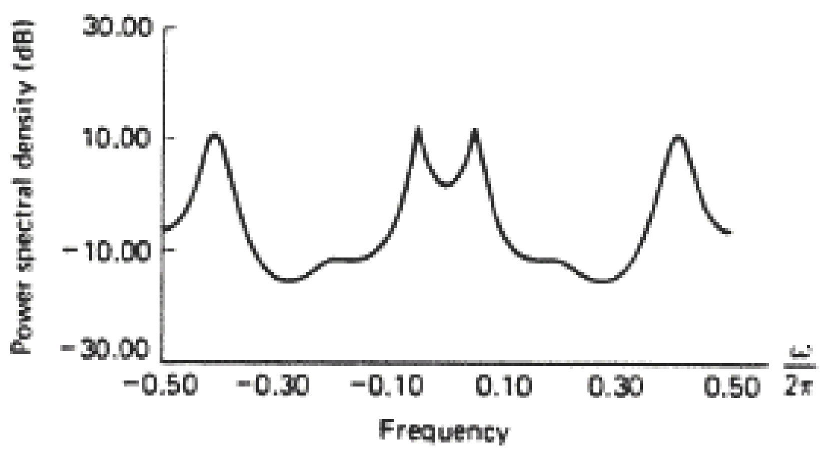



(b) Periodogram

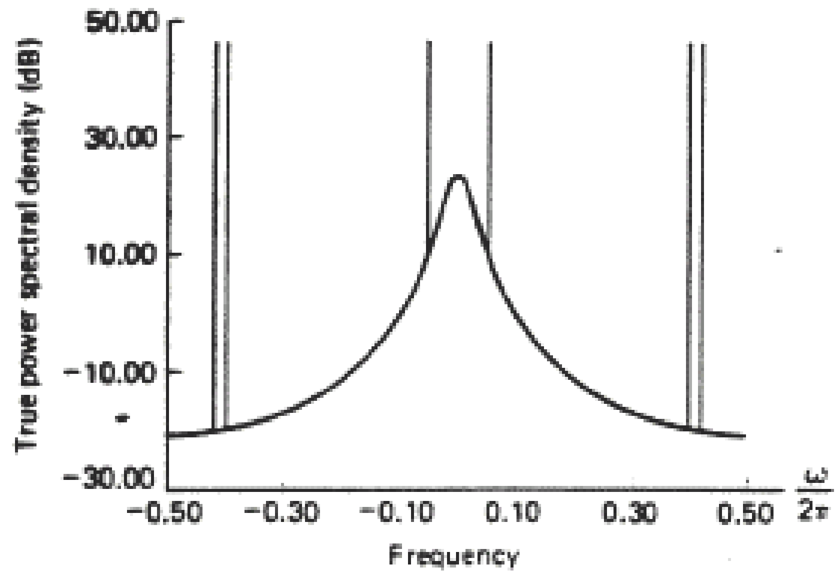
triangle window
M = 10




(c) Blackman–Tukey




(d) Minimum variance spectral estimator

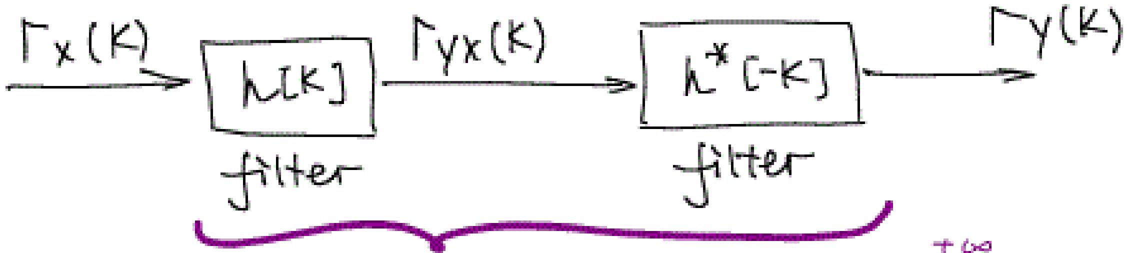



true p.s.d.

(a)

# ***Reference***

# Recall:  Filtering a Random Process


$$r_h[k] = h[k] * h^*[-k] = \sum_{l=-\infty}^{+\infty} h[l]\, h^*[k+l]$$

In terms of ZT:

$$P_y(z) = P_x(z)\, H(z)\, H^*(1/z^*) \quad \leftarrow h^*[-k]$$

$$\underset{z=e^{j\omega}}{\Rightarrow} P_y(\omega) = P_x(\omega)\, H(\omega)\, H^*(\omega) = P_x(\omega)\, |H(\omega)|^2.$$

## ***Chi-Squared Distribution***

If $x[n] \sim$ iid $N(0,1)$ for $n=0,1,\ldots N-1$, and

$$y = \sum_{n=0}^{N-1} x^2[n],$$

then $y$ follows chi-squared distribution of degree $N$, i.e. $y \sim \chi_N^2$

and $E[y] = N$, $Var(y) = 2N$

# Chi-Squared Distribution (cont'd)

p.d.f. of $y \sim \chi_N^2$ :

$$p(y) = \begin{cases} \frac{1}{2^{N/2}\Gamma(N/2)} y^{\frac{N}{2}-1} e^{-\frac{y}{2}} & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases}$$

where $\Gamma(\cdot)$ is the gamma integral

$$\Gamma(x+1) = \int_0^\infty y^x e^{-y} dy \quad \text{for } x > -1.$$

Note if $x$ is an integer, $\Gamma(n+1) = n\Gamma(n) = n!$

# ***Periodogram of White Gaussian Process***

For $f_K = K/N$, it can be shown that

$$\begin{cases} \dfrac{2\hat{P}_{PER}(f_K)}{P(f_K)} \sim \chi_2^2 & \text{for } K = 1, 2, \cdots \frac{N}{2}-1; \\ \dfrac{\hat{P}_{PER}(f_K)}{P(f_K)} \sim \chi_1^2 & \text{for } K = 0, \frac{N}{2} \end{cases}$$

$$\Rightarrow \quad E[\hat{P}_{PER}(f_K)] = P(f_K), \quad K = 0, 1, \cdots N/2$$

$$Var[\hat{P}_{PER}(f_K)] = \begin{cases} P^2(f_K), & K = 1, \cdots \frac{N}{2}-1 \\ 2P^2(f_K), & K = 0, \frac{N}{2} \end{cases}$$

See proof in Appendix 2.1 in Lim-Oppenheim Book:
- Basic idea is to examine the distribution of real and imaginary part of the DFT, and take the magnitude

Stamp image
from USPS web


# Happy Thanksgivings!