Statistical Signal Processing

7. Classic Methods for Spectrum Estimation

Electrical & Computer Engineering

North Carolina State University

Acknowledgment: ECE792-41 slides were adapted from ENEE630 slides developed by Profs. K.J. Ray Liu and Min Wu at the University of Maryland, College Park. Contact: chauwai.wong@ncsu.edu.

Summary of Readings

Overview   Haykin 1.16, 1.10

7. Non-parametric method

Hayes 8.1; 8.2 (8.2.3, 8.2.5); 8.3

8. Parametric method

Hayes 8.5, 4.7; 8.4

9. 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 w.s.s. 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
  ⇔ estimating autocorrelation at all lags

Spectral Estimation: Challenges

- A w.s.s process is infinitely long. (Why?) When a limited amount of observation data is 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

- 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

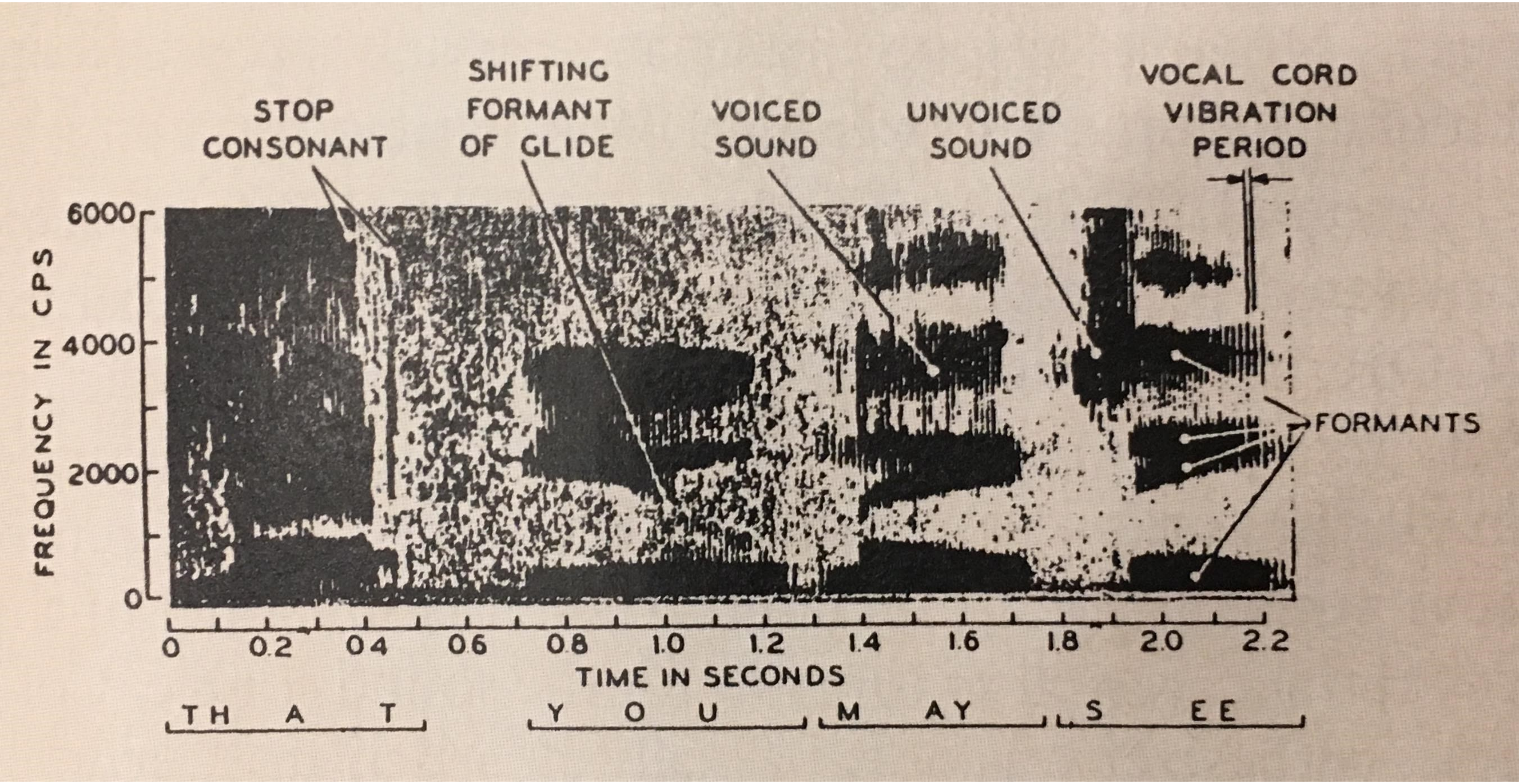
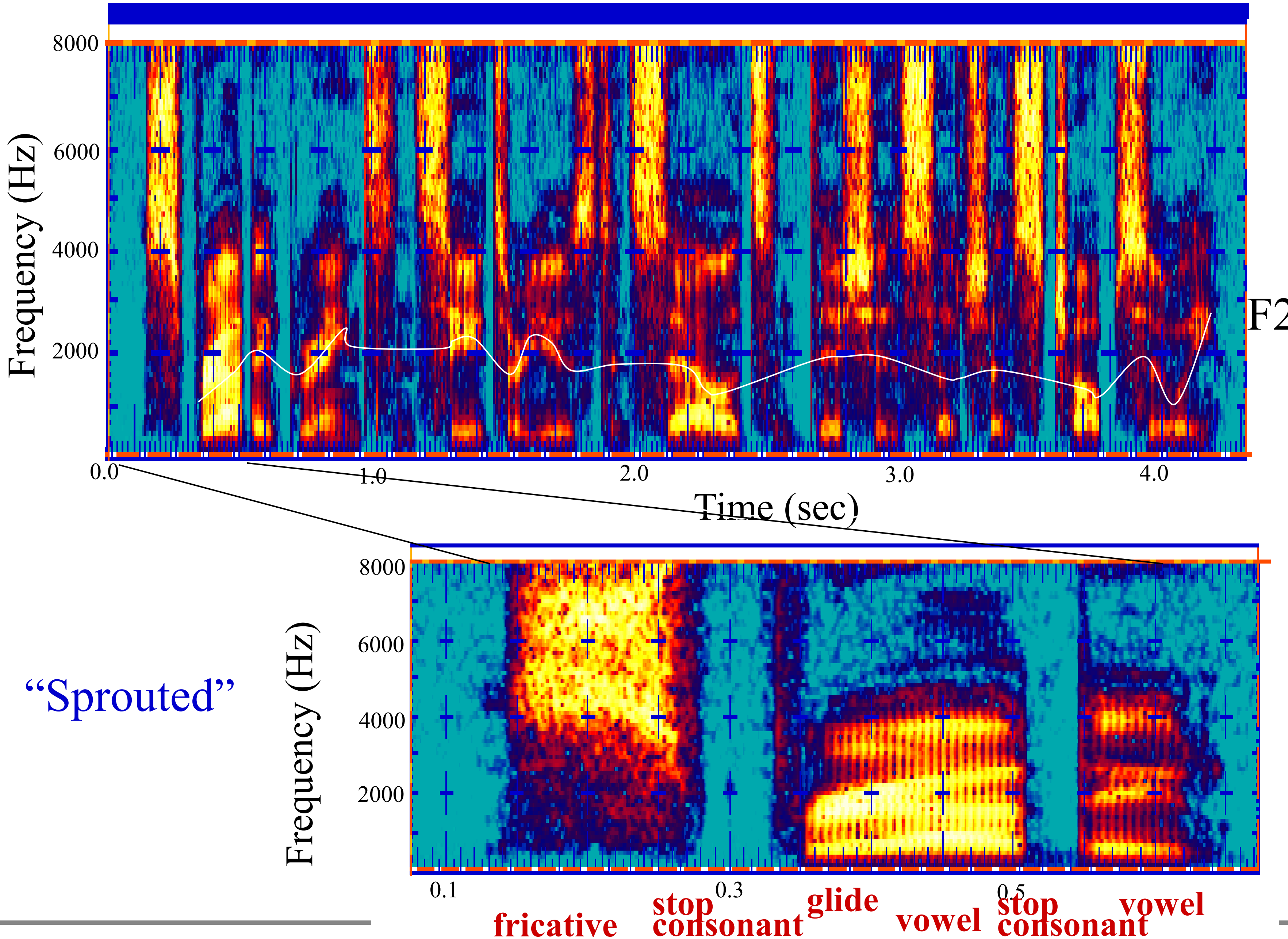


Figure 3 of SPM May'98 Speech Survey

“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

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

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

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 $|X(\omega)|^2$

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

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

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

Note: for each frequency f, $\hat{P}_M(f)$ is a random variable

Ensemble Average of $\hat{P}_M(f)$

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

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

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

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

- Now, what if *M* goes to infinity?

P.S.D. and Ensemble Fourier Magnitude

If the autocorrelation function decays fast enough s.t.

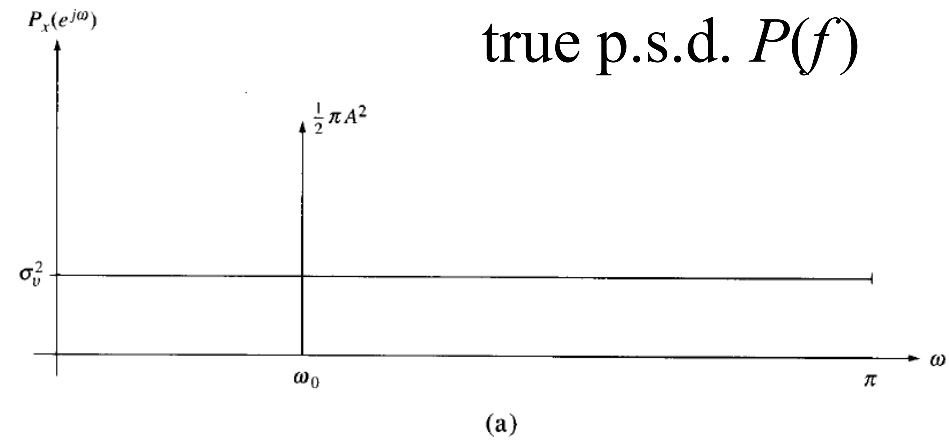
$$\sum_{k=-\infty}^{\infty} |k| r(k) < \infty \quad (\text{i.e., } r(k) \to 0 \text{ rapidly for } k \uparrow)$$

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

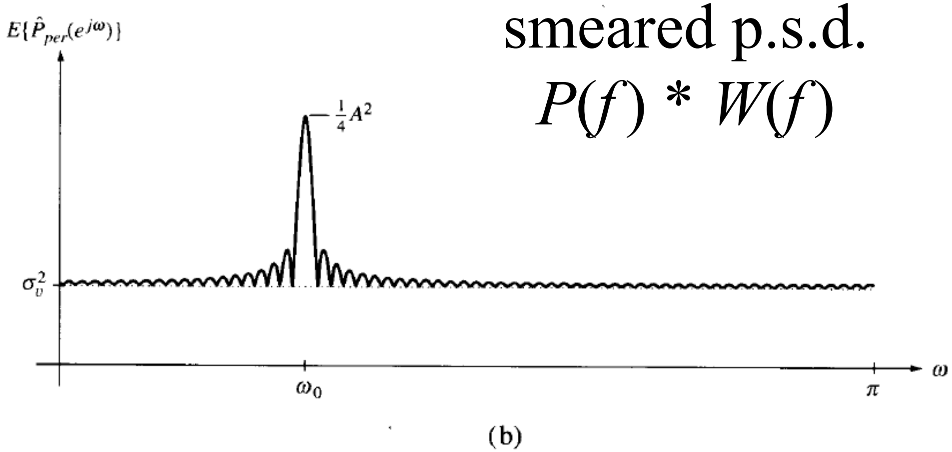
p.s.d.

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

Smeared P.S.D. for Finite Length Data



true p.s.d. $P(f)$



smeared p.s.d. $P(f) * W(f)$

Figure 8.5 *(a) The power spectrum of a single sinusoid in white noise and (b) the expected value of the periodogram.*

[Hayes Fig. 8.5]

Recall $E[\hat{P}_M(f)]$

$$= \sum_{k=-M}^{M} w(k) r(k) e^{-j2\pi fk}$$

$$= P(f) * W(f)$$

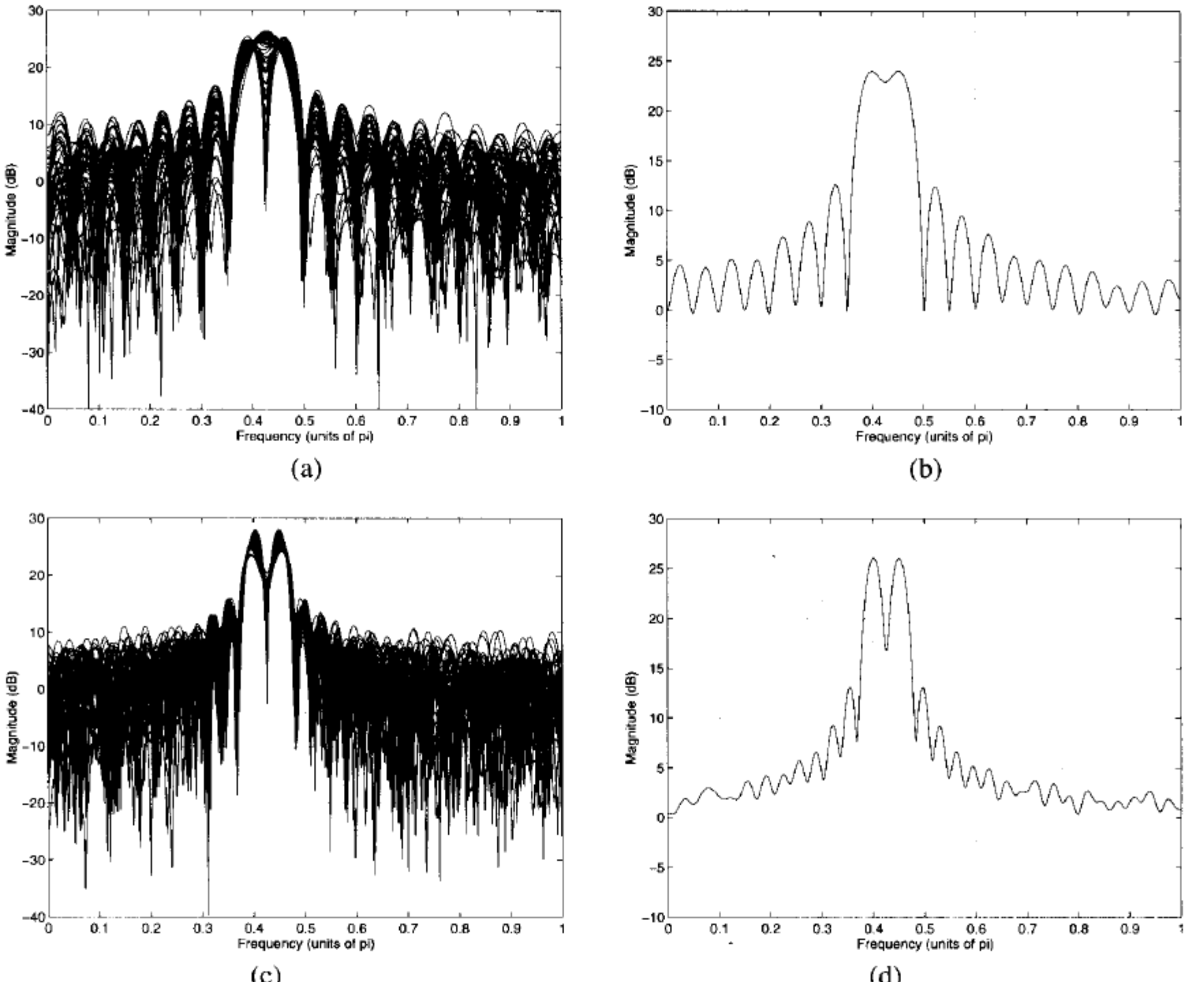
where

$$w(k) = 1 - \frac{|k|}{2M+1}$$

$$W(f) = \frac{\sin[(M+1)\omega/2]}{2\pi(M+1)\sin(\omega/2)}$$

Frequency Resolution Improves as N increases

50 realizations overlaid   averaged



[Hayes Fig. 8.8]

Frequency resolution: O(1/N)

Signal length N = 40

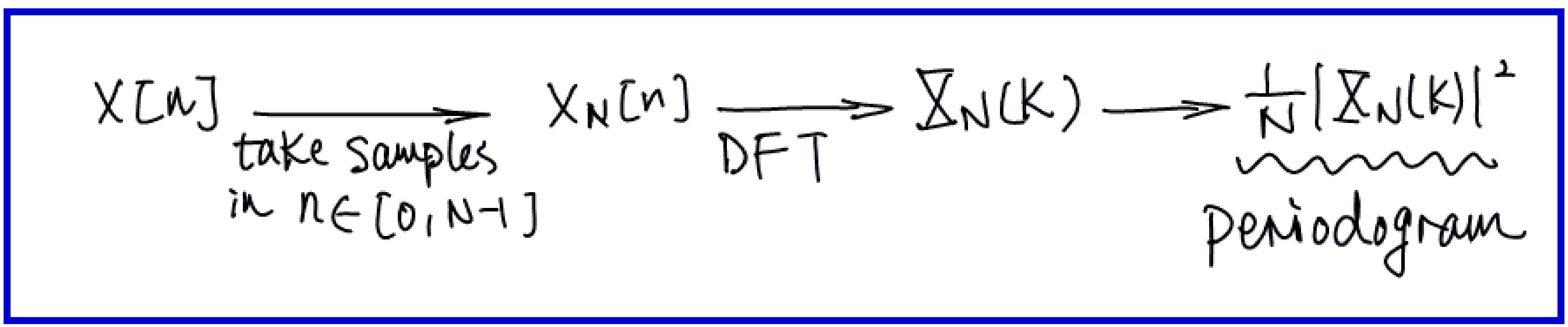
Signal length N = 64

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

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



An Equivalent Expression of Periodogram

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

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

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

– 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

– Autocorrelation sequence is zeroed out for $|k| \geq N$.

Exercise: Prove using the definition of the periodogram estimator.

(2) Filter Bank Interpretation of Periodogram

For a particular frequency of $f_{0:}$

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

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

where

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

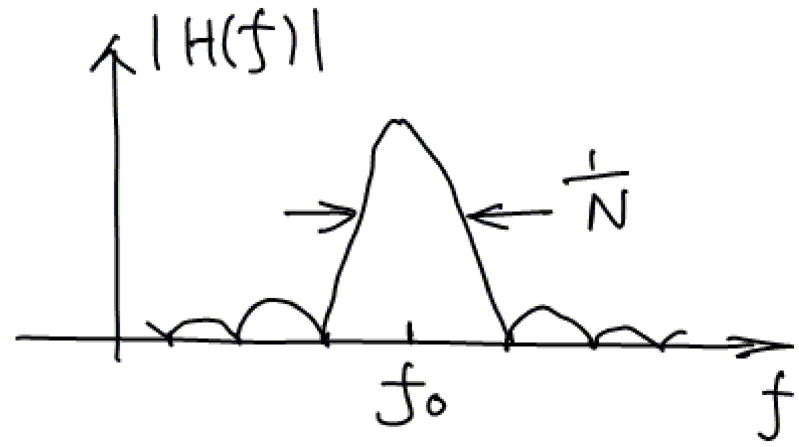
– Impulse response of the filter $h[n]$: a windowed version of a complex exponential

Frequency Response of h[n]

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

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

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



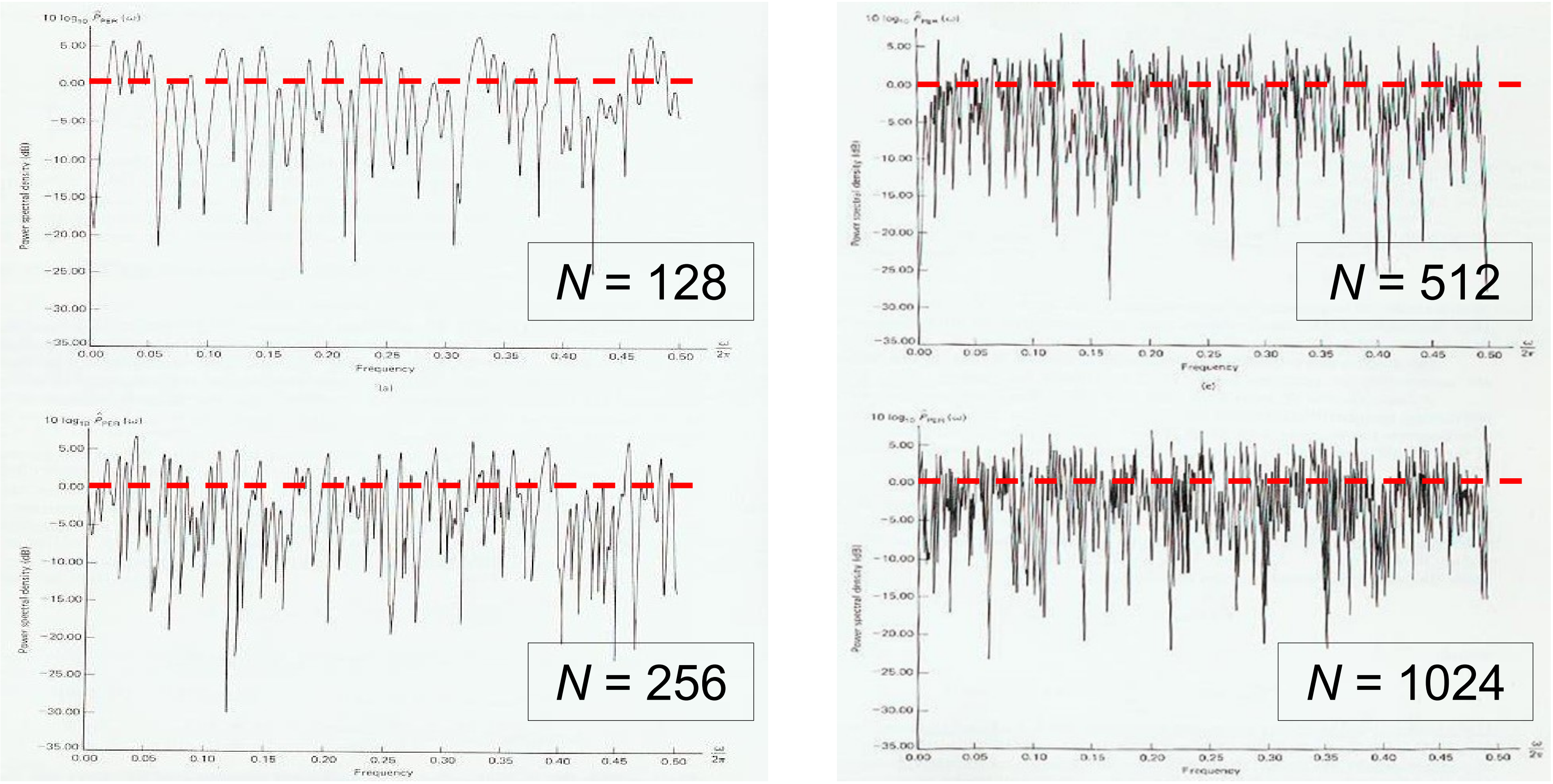
Periodogram: Filter Bank Perspective

- Can view the periodogram as an estimator of power spectrum 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

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

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 estimator does not decrease with increasing *N*

➔ periodogram is an inconsistent 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

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

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

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

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

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

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

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

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

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

and the $N = KL$ data points are arranged into K sets of length L:

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

Performance of Averaged Periodogram

– Assume the K sets of data records are mutually uncorrelated.

– For a white Gaussian input signal, $\hat{P}_{\text{AVPER}}^{(m)}(f), m = 0, \ldots, L-1$ are i.i.d., and one can verify that

$$\text{Var}[\hat{P}_{\text{AVPER}}(f_i)] =$$

$$\begin{cases} \frac{1}{K} P^2(f_i), & i = 1, 2, \cdots, \frac{L}{2} - 1, \\ \frac{2}{K} P^2(f_i), & i = 0, \frac{L}{2}, \end{cases}$$

where $f_i = i/L$.

– If L is fixed, K and N are allowed to go to infinity, then $\hat{P}_{\text{AVPER}}(f)$ is a consistent estimator.

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*):

$$x_m[n] = x[n + mL], \qquad n = 0, 1, \ldots, L-1$$
$$m = 0, 1, \ldots, K-1$$

- 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 Bartlett's method

Averaged Periodogram for Fixed Data Size

- Given a data record of fixed length *N*, will the result continue improving if we segment it into more and more subrecords?

We examine for a real-valued stationary process:

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

identical stat. mean for all *m*

Note

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

where

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

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

Mean of Averaged Periodogram

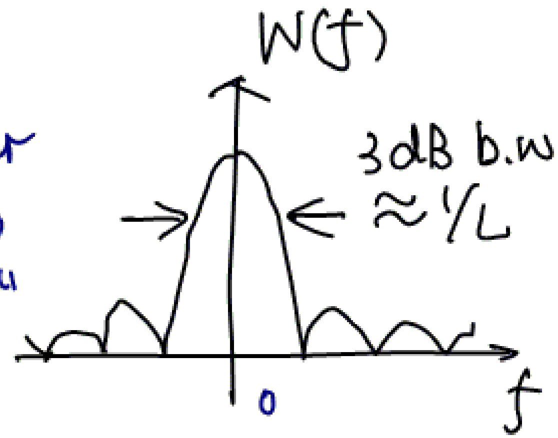
$$\Rightarrow E[\hat{r}^{(0)}(l)] = \underbrace{\left(1 - \frac{|l|}{L}\right)}_{\triangleq\, w[l]} r(l) \quad \text{for } |l| \le L-1$$

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

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

"triangular (Barlett) window"

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



Statistical Properties of Averaged Periodogram

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

multiplication in time

↓

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

convolution in frequency

$$\neq P(f)$$

- Biased estimator (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.
- ***Asymptotically*** unbiased as $L \to \infty$
  - To avoid the smearing, the window length L must be large enough so that the narrowest peak in $P(f)$ can be resolved
- Fixing $N = KL$, the choice of K leads to 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

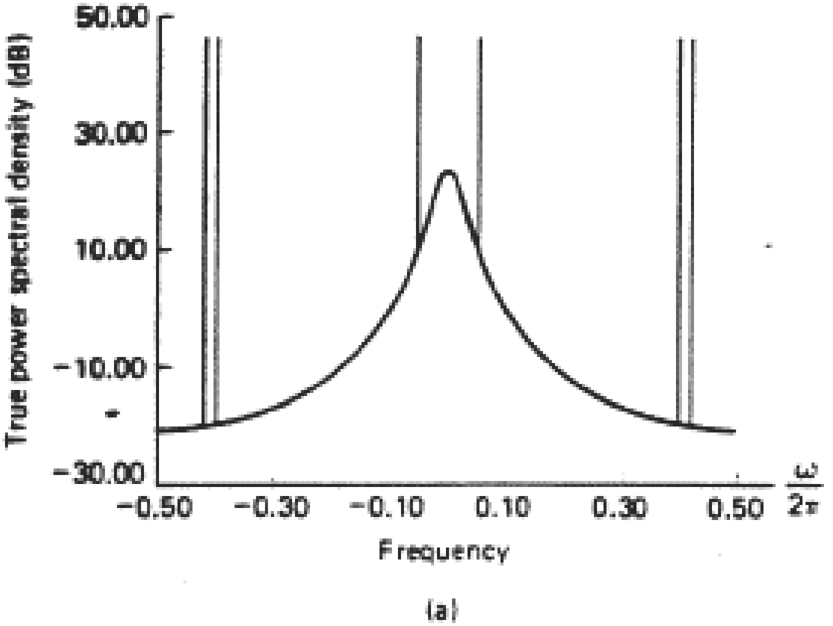
    (windowing in time gives smearing/smoothing in freq. domain)

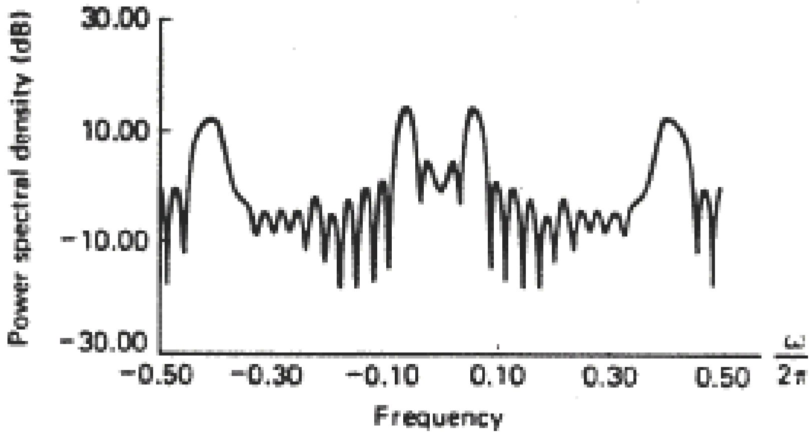
    unbiased for white Gaussian (flat spectrum)
- Averaged periodogram
  - Reduce variance by averaging *K* sets of data record of length *L* each
  - Small *L* increases smearing/smoothing in p.s.d. estimate thus higher bias ➔ *equiv. to triangular windowing to autocorrelation sequence*
- Windowed periodogram: 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 $z[n] = -a_1 z[n-1] + v[n], a_1 = -0.85, \sigma_v^2 = 0.1,$
    $\omega_1/2\pi = 0.05, \omega_2/2\pi = 0.40, \omega_3/2\pi = 0.42.$
  - $N = 32$ data points are available
    ➔ periodogram resolution $f = 1/32$
- Examine typical characteristics of various non-parametric spectral estimators

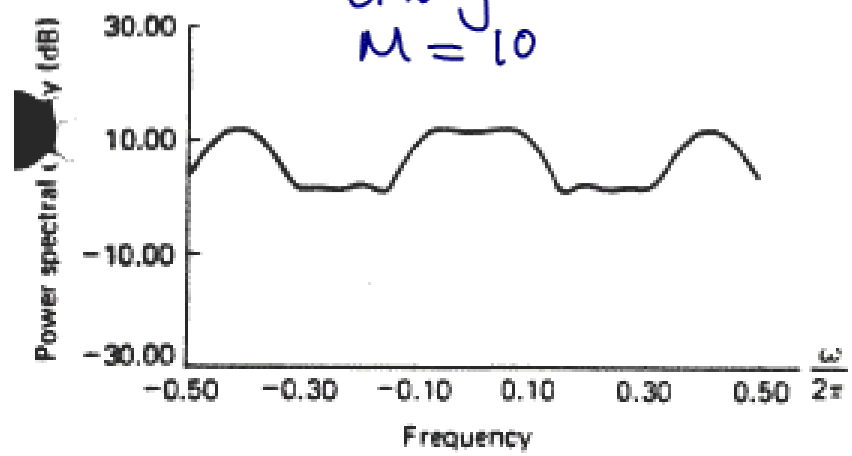
(Fig.2.17 from Lim/Oppenheim book)



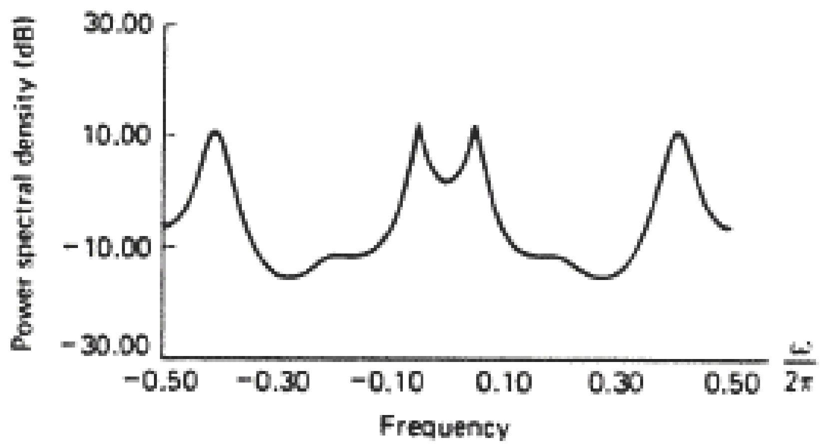


(b) Periodogram

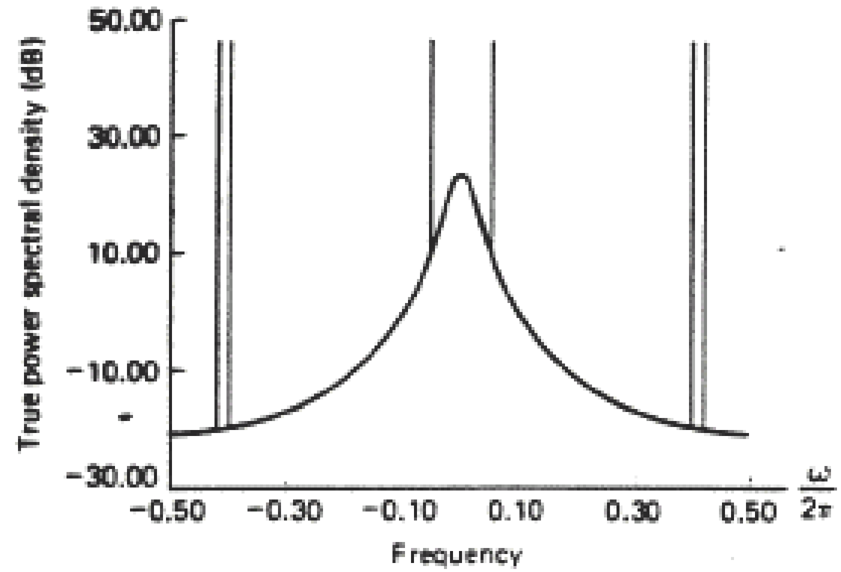
triangle window
M = 10



(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 $\hat{r}(k)$

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

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

  – 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)

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

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

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

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

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

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

1. 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

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

3. 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

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

subject to $H(f_0) = 1$

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

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

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

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

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

Equiv. to filtering $r(k)$ with $h(k) \otimes h^*(-k)$ and evaluating at output time $k = 0$

Matrix-Vector Form of MVSE Formulation

Define

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

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

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

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

Thus the problem becomes

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

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

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

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

and $\underline{h}^H \underline{e} = 1$

Solution to MVSE

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

$$\begin{cases} \nabla_{\lambda^*} \text{ or } \nabla_{\lambda} J = 0 \Rightarrow \underline{h}^{\mathrm{H}} \underline{e} = 1 \quad (*) \\ \nabla_{\underline{h}^*} \text{ or } \nabla_{\underline{h}} J = 0 \Rightarrow R^T \underline{h} - \lambda \underline{e} = 0 \Rightarrow \underline{h} = \lambda (R^T)^{-1} \underline{e} \quad (**) \end{cases}$$

Bring (∗∗) into (∗):

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

Filter's output power:

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

$$= \lambda$$

The optimal filter and its output power:

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

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

MVSE: Summary

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

Thus the MVSE is

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

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

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

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

- MVSE is a data adaptive estimator and provides improved resolution and reduced variance over periodogram
  - Also referred to as "High-Resolution Spectral Estimator"
  - Doesn't assume a particular underlying model for the data

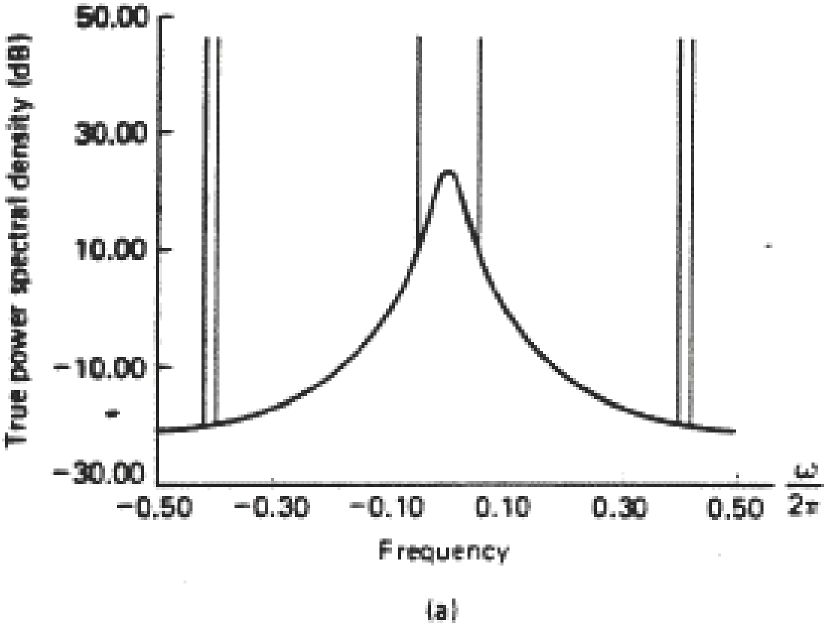
MVSE vs. Periodogram

- MVSE is a data adaptive estimator and provides improved resolution and reduced variance over periodogram

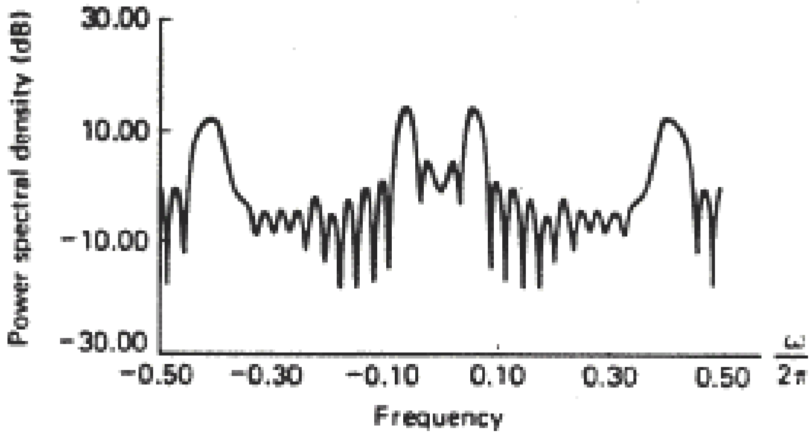
	Periodogram	**MVSE**
Equivalent Bandpass Filter $\underline{h}$	$\underline{e}$ Filter is "universal" data-independent	$$\frac{\left(R^T\right)^{-1}}{\underline{e}^H \left(R^T\right)^{-1} \underline{e}} \underline{e}$$ Filter adapts to observation data via *R*
Equivalent spectrum estimate $\hat{P}(f)$	$q \cdot \underline{e}^H \hat{R}^T \underline{e}$	$$\frac{q}{\underline{e}^H \left(\hat{R}^T\right)^{-1} \underline{e}}$$

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 $z[n] = -a_1 z[n-1] + v[n], a_1 = -0.85, \sigma_v^2 = 0.1,$
    $\omega_1/2\pi = 0.05, \omega_2/2\pi = 0.40, \omega_3/2\pi = 0.42.$
  - $N = 32$ data points are available
    ➔ periodogram resolution $f = 1/32$
- Examine typical characteristics of various non-parametric spectral estimators

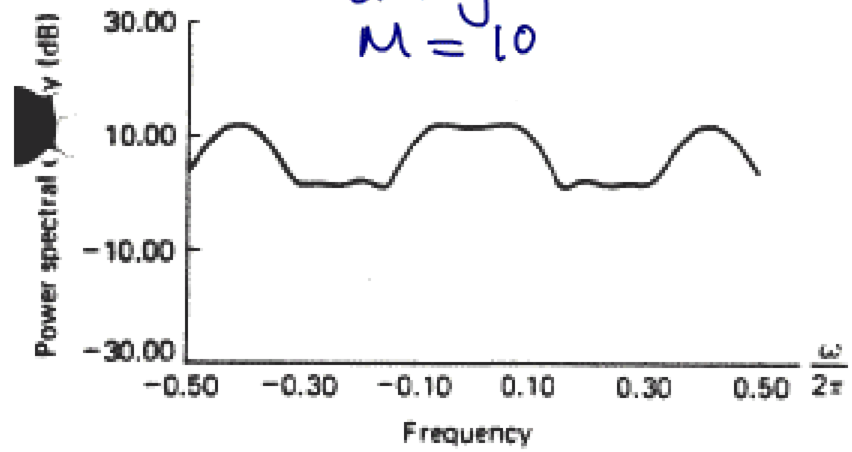


(Fig.2.17 from Lim/Oppenheim book)

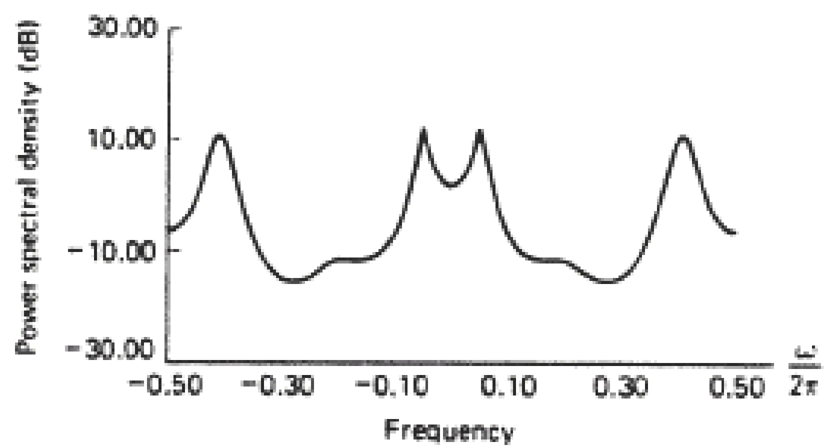


(b) Periodogram

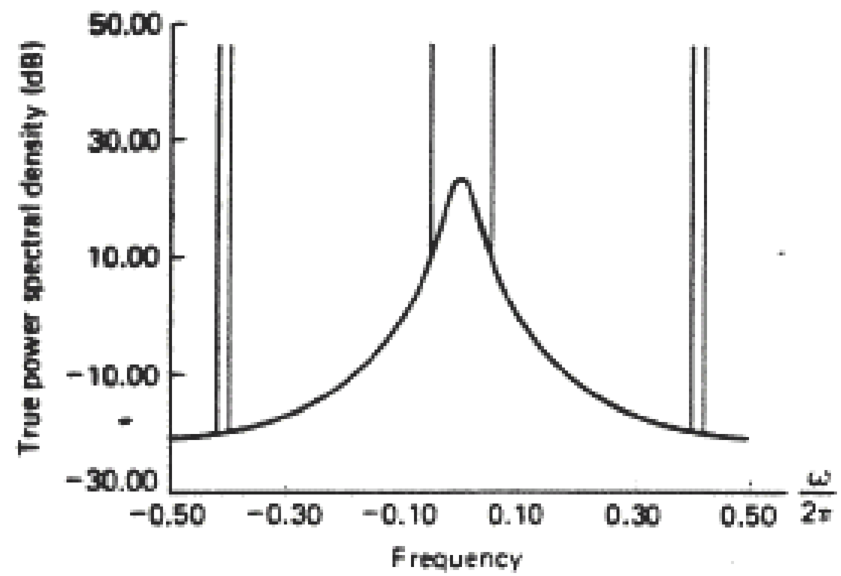
triangle window
M = 10



(c) Blackman-Tukey



(d) Minimum variance spectral estimator



true p.s.d.

(a)

Ref. on Derivative and Gradient Operators for Complex-Variable Functions

Ref: D.H. Brandwood, “A complex gradient operator and its application in adaptive array theory,” in IEE Proc., vol. 130, Parts F and H, no.1, Feb. 1983.

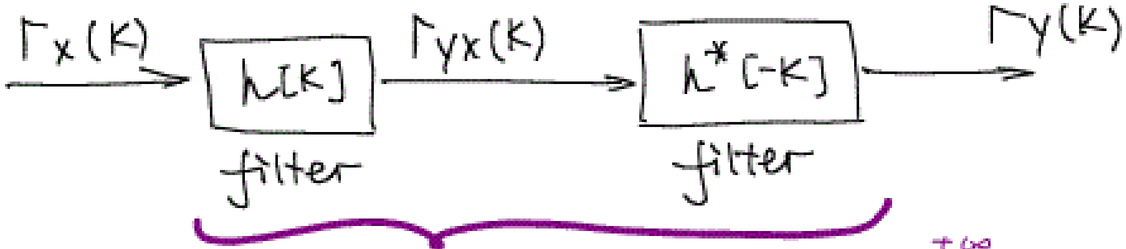
(downloadable from IEEEXplore)

– Solving constrained optimization with real-valued objective function of complex variables, subject to constraint function of complex variables

As seen in minimum variance spectral estimation and other array/statistical signal processing context.

Reference

Recall: Filtering a Random Process



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

In terms of ZT:

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

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

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