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

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

- 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

- 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

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

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

\[
P_e(f) = \sum_{k=-\infty}^{\infty} r(k) e^{-j2\pi fk}
\]

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

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

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

\[
P(f) = \sum_{k=-\infty}^{\infty} r(k)e^{-j2\pi fk}
\]

\(-\frac{1}{2} \leq f \leq \frac{1}{2}\)

(or \(\omega = 2\pi f : -\pi \leq \omega \leq \pi\))

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 $|X(\omega)|^2$

Consider

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

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

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

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

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

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.

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

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

... ... ... ... ...
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

\[ \lim_{M \to \infty} E[P_M(f)] = \sum_{k=-\infty}^{\infty} r(k)e^{-j2\pi fk} = P(f) \quad \text{p.s.d.} \]

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], \ldots, x[N-1]\}, the periodogram is defined as

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

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

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

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

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 $r(k)$

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

where

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

$\hat{r}(-k) = \hat{r}(k)$ 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 $f_0$:

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

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

where

$$h[n] = \text{Impulse response of the filter h[n]: a windowed version of a complex exponential}$$

Nonparametric spectral estimation [23]
(2) Filter Bank Interpretation of Periodogram

For a particular frequency of $f_0$:

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

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

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]**

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

sinc-like function centered at \( f_0 \):

![Plot of |H(f)|](image)
**Frequency Response of h[n]**

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

Sinc-like function centered at \( f_0 \):

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

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

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

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
  \(\Rightarrow\) periodogram is not a consistent estimator
(3) How Good is Periodogram for Spectral Estimation?

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

- Estimation: Tradeoff between bias and variance

\[
E(\hat{\theta}) \neq \theta
\]

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

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

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

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

\[
P_{AV\,PER}(f) = \frac{1}{K} \sum_{m=0}^{K-1} P_{PER}(f)
\]

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

And the K sets of data records are
\[
\{x_0[0], \ldots, x_0[L-1]; x_1[n], 0 \leq n \leq L-1; \ldots\}
\]
\[
\{x_{K-1}[n-1], 0 \leq n \leq L-1\}
\]
Performance of Averaged Periodogram

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

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

\[ \Rightarrow \text{Var}[\hat{P}_{AVPER}(f)] = \frac{1}{K} P^2(f_i) \]
Performance of Averaged Periodogram

- If K sets of data records are uncorrelated with each other, we have:

\[ (f_i = i/L) \]

\[ \hat{P}_\text{PER}^{(m)}(f) \quad \text{i.i.d. (m=0,1, \ldots, L-1)} \] for white Gaussian process

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

i.e., \( K \uparrow \rightarrow \text{Var} \downarrow \), and \( \text{Var} \rightarrow 0 \) for \( K \rightarrow \infty \)

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. $N=KL$)
  
  \[ x_m[n] = x[n + mL], \quad n = 0, 1, ..., L - 1 \]
  \[ m = 0, 1, ..., K - 1 \]

- 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:

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

identical distribution 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

\[ \hat{r}^{(0)}(l) = \frac{1}{L} \sum_{n=0}^{L-1-|l|} x[n]x[n + |l|] \]
Mean of Averaged Periodogram

\[
W[k] = \begin{cases} 
1 - \frac{K}{L} & \text{for } |K| \leq L-1 \\
0 & \text{o.w.}
\end{cases}
\]

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

"triangular (Barlett) window"

3dB b.w. \( \approx \frac{1}{L} \)
Mean of Averaged Periodogram

\[
E[\hat{P}_n(t)] = (1 - \frac{|l|}{L}) \lambda(l) \quad \text{for } |l| \leq L - 1
\]

\[
\Delta \triangleq \mathbf{W}(l)
\]

\[
E[\hat{P}_{\text{AvPER}}(f)] = \sum_{l=-L-1}^{L-1} \mathbf{W}(l) \lambda(l) e^{-j2\pi fl}
\]

\[
\mathbf{W}(k) = \begin{cases} 
1 - \frac{|k|}{L} & \text{for } |k| \leq L - 1 \\
0 & \text{otherwise}
\end{cases}
\]

\[
\mathbf{W}(f) = \frac{1}{L} \left( \frac{\sin \pi f L}{\sin \pi f} \right)^2
\]

"triangular \hspace{1cm} \text{(Bartlett) window}"

3dB b.w. \quad \approx \frac{1}{L}
Mean of Averaged Periodogram (cont’d)

\[ E[\hat{P}_{\text{AV PER}}(f)] = \text{DTFT}\{w[k]r(k)\} \]

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

\[ \neq P(f) \]

- **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)

\[ E[\hat{P}_{AVPER}(f)] = DTFT[\{w[k]r(k)\}]_f \]

\[ = \int_{-1/2}^{1/2} W(f - \eta)P(\eta)d\eta \]

\[ \neq P(f) \]

- **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 \( L \to \infty \)**
  - To avoid the smearing, the window length \( L \) must be large enough so that the narrowest peak in \( P(f) \) can be resolved

- **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

  \[ \Rightarrow \text{periodogram resolution } f = 1/32 \]

- Examine typical characteristics of various non-parametric spectral estimators

  (Fig.2.17 from Lim/Oppenheim book)
UMD ENEE630 Advanced Signal Processing (ver.1111)

Nonparametric spectral estimation

Triangle window

M = 10

True p.s.d.
3.1.3 Periodogram with Windowing

- **Review and Motivation**

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

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

where

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

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

\[
\hat{P}_{\text{win}}(f) = \sum_{K=-(N-1)}^{N-1} W(K) \hat{r}(K) e^{-j2\pi f K}
\]

where \( W(K) \) is a “lag window” with properties of:

- 1. \( 0 \leq W(K) \leq W[0] = 1 \)
- 2. \( W[-K] = W[K] \) symmetric
- 3. \( W[K] = 0 \) for \(|K| > M\) where \( M \leq N-1 \)
- 4. \( W(f) \) must be chosen to ensure \( \hat{P}_{\text{win}}(f) \geq 0 \)

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

\( w(0)=1 \) preserves variance \( r(0) \)
**Common Lag Windows**

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Fourier Transform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectangular</td>
<td>( w(k) = \begin{cases} 1, &amp;</td>
<td>k</td>
</tr>
<tr>
<td>Bartlett</td>
<td>( w(k) = \begin{cases} 1 - \frac{</td>
<td>k</td>
</tr>
<tr>
<td>Hanning</td>
<td>( w(k) = \begin{cases} \frac{1}{2} + \frac{1}{2} \cos \frac{\pi k}{M}, &amp;</td>
<td>k</td>
</tr>
<tr>
<td>Hamming</td>
<td>( w(k) = \begin{cases} 0.54 + 0.46 \cos \frac{\pi k}{M}, &amp;</td>
<td>k</td>
</tr>
<tr>
<td>Parzen</td>
<td>( w(k) = \begin{cases} 2 \left( 1 - \frac{</td>
<td>k</td>
</tr>
</tbody>
</table>

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 \(1/(N-k)\)?

**Biased** (low variance) \( \hat{r}_1(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x[n+k] x^*[n] \) \( \hat{r}_1(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x[n+k] x^*[n] \)

**Unbiased** (may not non-neg. definite) \( \hat{r}_2(k) = \frac{1}{N-k} \sum_{n=0}^{N-1-k} x[n+k] x^*[n] \)

\[ \mathbb{E}(\hat{r}_i(k)) = \]

- Hints on showing the non-negative definiteness: using \( r_1(k) \)
to construct correlation matrix

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

UMD ENEE630 Advanced Signal Processing (ver.1111)    Nonparametric spectral estimation [47]
**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) ?

  **Biased** (low variance) : $\hat{r}_1(k) = \frac{1}{N} \sum_{n=0}^{N-1-K} X[n+k] X^*[n]$  

  Bias: $E(\hat{r}_1(k)) = \frac{N-K}{N} r(k)$

  **Unbiased** (may not non-neg. definite) : $\hat{r}_2(k) = \frac{1}{N-K} \sum_{n=0}^{N-1-K} X[n+k] X^*[n]$  

  Unbias: $E(\hat{r}_2(k)) = r(k)$

- Hints on showing the non-negative definiteness: using $r_1(k)$ to construct correlation matrix
- For $\hat{r}_2(k)$ : HW#8

\[
\hat{R}_N = X^H X, \text{ where } \quad X = \frac{1}{\sqrt{N}} \begin{bmatrix}
X(0) & 0 & 0 \\
X(1) & X(0) & 0 \\
\vdots & \vdots & \ddots & \ddots \\
X(N-1) & \cdots & X(0) & 0 \\
0 & \cdots & X(N-1) & 0 \\
\end{bmatrix}
\]
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

- 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

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

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:

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

Thus

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

Equiv. to filter r(k) with \( \{ h(k) \otimes h^*(-k) \} \) 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

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

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

Equiv. to filter $r(k)$ with \{ $h(k) \otimes h^*(-k)$ \} and evaluate at output time $k=0$
Matrix-Vector Form of MVSE Formulation

Define

\[
\mathbf{h}^* \triangleq \begin{bmatrix}
    h[0] \\
h[-1] \\
    \vdots \\
h[-(N-1)]
\end{bmatrix}
\]

\[
\Rightarrow \mathbf{p} = \mathbf{h}^H \mathbf{R}^{-1} \mathbf{h}
\]

\[
\mathbf{e} = \begin{bmatrix}
    e^{j2\pi f_0} \\
    \vdots \\
    e^{j2\pi (N-1)f_0}
\end{bmatrix}
\]

The constraint can be written in vector form as

\[
\mathbf{h}^H \mathbf{e} = 1
\]

\[
H(f_0)
\]
Matrix-Vector Form of MVSE Formulation

Define

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

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

\[ \Rightarrow \text{The constraint can be written in vector form as } h^H e = 1 \]

Thus the problem becomes

\[ \min_h h^H R^T h \quad \text{subject to } h^H e = 1 \]
Solution of MVSE

\[ J = h^H R^T h + \text{Re} \left[ 2\lambda (1 - h^H e) \right] \]

- 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

\[
\begin{align*}
\min_{\mathbf{h}, \lambda} J &= \mathbf{h}^H R^T \mathbf{h} + \lambda (1 - \mathbf{h}^H \mathbf{e}) + \left[ \lambda (1 - \mathbf{h}^H \mathbf{e}) \right]^* \\
&= \mathbf{h}^H R^T \mathbf{h} + \lambda (1 - \mathbf{h}^H \mathbf{e}) + \lambda^*(1 - \mathbf{e}^H \mathbf{h}) \\
\end{align*}
\]

\[
\begin{align*}
\nabla_{\mathbf{h}} J &= 0 \Rightarrow R^T \mathbf{h} - \lambda \mathbf{e} = 0 \\
\nabla_{\mathbf{h}} J &= 0 \Rightarrow \left( \mathbf{h}^H R^T \right)^T - \lambda^* \mathbf{e}^* = 0 \\
\Rightarrow \left( R^T \right)^H \mathbf{h} - \lambda \mathbf{e} = 0 \Rightarrow R^T \mathbf{h} - \lambda \mathbf{e} = 0
\end{align*}
\]

\[
\begin{align*}
\Rightarrow \mathbf{h} &= \lambda \left( R^T \right)^{-1} \mathbf{e} \quad \text{and} \quad \mathbf{h}^H \mathbf{e} = 1 \\
\Rightarrow \begin{cases} 
\lambda &= \frac{1}{\mathbf{e}^H \left( R^T \right)^{-1} \mathbf{e}} \\
\mathbf{h} &= \frac{\left( R^T \right)^{-1} \mathbf{e}}{\mathbf{e}^H \left( R^T \right)^{-1} \mathbf{e}} 
\end{cases}
\end{align*}
\]

UMD ENEE630 Advanced Signal Processing (ver.1111)
Solution of MVSE (cont’d)

The optimal filter:

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

It follows that

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

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

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

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

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

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

$e = \begin{bmatrix}
e_1 \\
e_2 \\
\vdots \\
e_p
\end{bmatrix}$

- 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
    \( \implies \) periodogram resolution \( f = 1/32 \)

- Examine typical characteristics of various non-parametric spectral estimators

  (Fig.2.17 from Lim/Oppenheim book)
Triangle window

$m = 10$

true p.s.d.
Reference
Recall: Filtering a Random Process

\[ \Gamma_x(K) \xrightarrow{\text{h}[K]} \Gamma_y(K) \xrightarrow{\text{h}^*[K]} \Gamma_y(K) \]

\[ \Gamma_h[K] = h[K] \ast h^*[-K] = \sum_{l=-\infty}^{\infty} h[l] h^*[K+l] \]

In terms of \( \mathbb{E} \): 
\[ P_y(\delta) = P_x(\delta) H(\delta) H^*(\delta^*) \]
\[ \Rightarrow P_y(\omega) = P_x(\omega) H(\omega) H^*(\omega) = P_x(\omega) |H(\omega)|^2. \]
Chi-Squared Distribution

If \( x[n] \sim \text{iid} \ N(0,1) \) 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. \( y \sim \chi^2_N \)

and \( E[y] = N \), \( \text{Var}(y) = 2N \)
Chi-Squared Distribution (cont’d)

P.d.f. of \( y \sim \chi^2_N \):

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

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

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

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

\[
\frac{2 \hat{P}_{\text{PER}}(f_k)}{P(f_k)} \sim X^2 \quad \text{for} \quad k = 1, 2, \ldots, \frac{N}{2} - 1, \\
\frac{\hat{P}_{\text{PER}}(f_k)}{P(f_k)} \sim X^1 \quad \text{for} \quad k = 0, \frac{N}{2}
\]

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

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

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
Happy Thanksgivings!