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@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

- **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

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

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

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

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

Note: for each frequency $f$, $P_M(f)$ 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

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

Thus
If the autocorrelation function decays fast enough s.t.

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

then

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

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

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

\[X[n] \xrightarrow{\text{take samples in } n \in [0, N-1]} X_N[n] \xrightarrow{\text{DFT}} \hat{X}_N(K) \xrightarrow{\text{periodogram}} \frac{1}{N} |\hat{X}_N(K)|^2\]
An Equivalent Expression of Periodogram

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

– 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 $\hat{r}(k)$

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

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

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

$$h[n] = \ldots$$

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

© 2003 by M.Wu
(2) Filter Bank Interpretation of Periodogram

For a particular frequency of $f_0$:

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

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

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 $f_0$: 

![Diagram of frequency response](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 \( \sim \) a set of bandpass filters

\[
\hat{P}_{\text{PER}}(f_0) = \left[ N \cdot \sum_{k=0}^{N-1} h[n-k]x[k] \right]_n^2
\]
**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
  \[ \text{periodogram is not a consistent estimator} \]
(3) How Good is Periodogram for Spectral Estimation?

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

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

where

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

And the K sets of data records are

\[
\{x_0[0], \ldots, x_0[L-1]; x_1[n], 0 \leq n \leq L-1; \ldots \}
\]

\[
\{x_{K-1}[n-1], 0 \leq n \leq L-1 \}
\]
\textbf{Performance of Averaged Periodogram}

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

\[
(f_i = i/L)
\]

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

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

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

\[ (f_i = i/L) \]

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

\[ \Rightarrow \text{Var} \left[ \hat{P}_{\text{AVPER}}(f) \right] = \frac{1}{K} P^2(f_i) \]

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

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

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

- 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 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:

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

Note

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

where

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

Note

Note
Mean of Averaged Periodogram

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

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

"triangular (Barlett) window"
Mean of Averaged Periodogram

\[ \Rightarrow E[\hat{P}^{(0)}(f)] = \left( 1 - \frac{|l|}{L} \right) \frac{1}{W(l)} \quad \text{for } |l| \leq L-1 \]

\[ \Rightarrow \quad E[\hat{P}_{\text{Aver}}(f)] = \sum_{l=-L}^{L-1} W(l) \frac{1}{W(l)} \cdot e^{-j2\pi fl} \]

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

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

\[ W(f) \]

- triangular window
- 3dB b.w. \approx \frac{L}{\pi f}
Mean of Averaged Periodogram (cont’d)

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

\[
= \int_{-\frac{3}{2}}^{\frac{1}{2}} W(f - \eta)P(\eta)d\eta
\]

\[
\neq P(f)
\]

- **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 (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

- **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
  \( \Rightarrow \) 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.
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

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

  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

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

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 \( \Leftrightarrow \) 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>
<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} \left( 1 + \frac{1}{2} \cos \frac{\pi k}{M} \right), &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)

Nonparametric spectral estimation [46]
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)  
  \[
  \hat{r}_1(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} X[n+k] X^*[n]
  \]

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

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

- For \( r_2(k) \): HW#8
Discussion: Estimate \( r(k) \) via Time Average

- Normalizing the sum of \((N-k)\) pairs
  
  \[ \hat{r}_1(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} X[n+k] X^*[n] \]
  
  \[ \hat{r}_2(k) = \frac{1}{N-k} \sum_{n=0}^{N-1-k} X[n+k] X^*[n] \]

  \[ \mathbb{E}(\hat{r}_1(k)) = \frac{N-K}{N} r(K) \]
  
  \[ \mathbb{E}(\hat{r}_2(k)) = r(K) \]

  Biased (low variance) vs. Unbiased (may not non-neg. definite)

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

- 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

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

(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:

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

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

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

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

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

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

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

⇒ The constraint can be written in vector form as \( h^H e = 1 \)

\[ \frac{1}{H(f_0)} \]
Matrix-Vector Form of MVSE Formulation

Define

\[ h^* \Delta \begin{bmatrix} h[0] \\ h[-1] \\ \vdots \\ h[-(N-1)] \end{bmatrix} \implies \rho = h^H R^T h \]

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

\[ H(f_0) \]

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

Thus the problem becomes

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

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

- 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

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

either \( \nabla_h J = 0 \Rightarrow R^T h - \lambda e = 0 \)

or \( \nabla_h J = 0 \Rightarrow \left( h^H R^T \right)^T - \lambda^* e^* = 0 \)

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

\( \Rightarrow h = \lambda (R^T)^{-1} e \)

and \( h^H e = 1 \)
Solution to MVSE

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

\[
\begin{aligned}
\nabla_{\lambda^*} J &= 0 \Rightarrow h^H e = 1 \quad (*) \\
\nabla_{h^*} J &= 0 \Rightarrow R^T h - \lambda e = 0 \Rightarrow h = \lambda (R^T)^{-1} e \\n\end{aligned}
\]

Bring (**) into (*):

\[
\lambda = \frac{1}{e^H (R^T)^{-1} e}
\]

Filter’s output power:

\[
\begin{aligned}
\rho &= h^H R^T h = h^H R^T (R^T)^{-1} e \lambda \\
&= \lambda \\
\end{aligned}
\]

The optimal filter and its output power:

\[
\begin{aligned}
\underline{h}_{MV} &= \frac{(R^T)^{-1}}{e^H (R^T)^{-1} e} e \\
\rho &= \frac{1}{e^H (R^T)^{-1} e}
\end{aligned}
\]
**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

\[ \hat{P}_{MV}(f) = \frac{q}{e^H (\hat{R}^T)^{-1} e} \]

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

\( \hat{R} \) is \( q \times q \) correlation matrix

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

- 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.

<table>
<thead>
<tr>
<th></th>
<th>Periodogram</th>
<th>MVSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equivalent Bandpass</td>
<td>$e$</td>
<td>$e - \frac{\left(R^T\right)^{-1} e}{e^H \left(R^T\right)^{-1} e}$</td>
</tr>
<tr>
<td>Filter is “universal” data-independent</td>
<td>$q \cdot e^H \hat{R}^T e$</td>
<td>$\frac{q}{e^H (\hat{R}^T)^{-1} e}$</td>
</tr>
<tr>
<td>Equivalent spectrum</td>
<td>$e$</td>
<td>Filter adapts to observation data via $R$</td>
</tr>
<tr>
<td>estimate $\hat{P}(f)$</td>
<td>$q \cdot e^H \hat{R}^T e$</td>
<td></td>
</tr>
</tbody>
</table>
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
    \[ \rightarrow \text{periodogram resolution } f = 1/32 \]

- Examine typical characteristics of various non-parametric spectral estimators

  (Fig.2.17 from Lim/Oppenheim book)
UMD ENEE630 Advanced Signal Processing (ver.1211)

nonparametric spectral estimation [63]
Ref. on Derivative and Gradient Operators for Complex-Variable Functions

(downloadable from IEEEXplorer)

– 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

\[
\begin{align*}
\text{W.S.S. process: } & \quad \{X[n]\} \\
\text{Filter: } & \quad \{h[n]\} \\
\text{Stable LTI filter: } & \quad \{y[n]\}
\end{align*}
\]

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

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

In terms of \( \mathcal{F} \):

\[
P_y(\beta) = P_x(\beta) H(\beta) H^* (\beta^*)
\]

\[
\Rightarrow P_y(w) = P_x(w) |H(w)|^2
\]

\[
\beta = e^{j\omega} = e^{jw}
\]
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 \( \mathbb{E}[y] = N, \ Var(y) = 2N \)
Chi-Squared Distribution (cont’d)

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

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

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

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

Note if \( x \) is an integer, \( \Gamma(x+1) = n! \Gamma(n) = n! \).
Periodogram of White Gaussian Process

For $f_k = k/N$, it can be shown that

$$
\frac{2\hat{P}_{\text{PER}}(f_k)}{P(f_k)} \sim \chi^2_{\nu} \quad \text{for } k=1,2,\ldots,\frac{N}{2}-1,
$$

$$
\frac{\hat{P}_{\text{PER}}(f_k)}{P(f_k)} \sim \chi^2_{\nu} \quad \text{for } k=0, \frac{N}{2}
$$

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

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

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
Preview & Warm-up
**Model or Not?**

- **Implicit assumption by classical methods**
  - Classical methods use Fourier transform on either windowed data or windowed autocorrelation function (ACF)
  - **Implicitly assume** the unobserved data or ACF outside the window are zero => not true in reality
  - Consequence of windowing: smeared spectral estimate (leading to low resolution)

- **If prior knowledge about the process is available**
  - Can use **prior knowledge** and select a good model to approximate the process
  - Usually need to estimate **fewer model parameters** (than non-parametric approaches) using the limited data points we have
  - The model may allow to **better describe the process outside the window** (instead of assuming zeros)
General Procedure of Parametric Methods

- Select a model (based on prior knowledge)
- Estimate the parameters of the assumed model
- Obtain the spectral estimate implied by the model (with the estimated parameters)
Spectral Estimation using AR, MA, ARMA Models

● Physical insight: the process is generated/approximated by filtering white noise with an LTI filter of rational transfer func H(z)

● Use observed data to estimate a few lags of r(k)
  – Larger lags of r(k) can be implicitly extrapolated by the model

● Relation between r(k) and filter parameters \{a_k\} and \{b_k\}
  – PARAMETER EQUATIONS from Section 2.1.2(6) \(\rightarrow\) review this
  – Solve the parameter equations to obtain filter parameters
  – Use the p.s.d. implied by the model as our spectral estimate

● Deal with nonlinear parameter equations
  – Try to “convert” or relate them to AR models that has linear equations
**Review: Parameter Equations**

Yule-Walker equations (for AR process)

\[
\gamma_x[k] = \begin{cases} 
- \sum_{l=1}^{p} a[l] \gamma_x[-l] + \sigma^2, & \text{for } k = 0 \\
- \sum_{l=1}^{p} a[l] \gamma_x[k-l], & \text{for } k \geq 1
\end{cases}
\]

ARMA model

\[
\gamma_x[k] = \begin{cases} 
- \sum_{l=1}^{p} a[l] \gamma_x[k-l] + \sigma^2 \sum_{l=0}^{q} b[l] \gamma_x[k-l], & k = 0, 1, \ldots, q \\
- \sum_{l=1}^{p} a[l] \gamma_x[k-l], & k \geq q+1
\end{cases}
\]

MA model

\[
\gamma_x[k] = \begin{cases} 
0, & \text{for } k = 0, 1, \ldots, q \\
\sigma^2 \sum_{l=0}^{q} b[l] \gamma_x[k-l], & \text{for } k \geq q+1
\end{cases}
\]
**Spectrum Estimation with AR Modeling**

- Use Levinson-Durbin recursion and solve for

\[
\begin{bmatrix}
\hat{r}(0) & \hat{r}(1) & \cdots & \hat{r}(L-1) \\
\hat{r}(1) & \hat{r}(0) & \cdots & \hat{r}(L-2) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{r}(L-1) & \cdots & \hat{r}(0) & \hat{r}(L-2)
\end{bmatrix}
\begin{bmatrix}
\hat{a}_1 \\
\hat{a}_2 \\
\vdots \\
\hat{a}_L
\end{bmatrix} = - \begin{bmatrix}
\hat{r}(1) \\
\hat{r}(2) \\
\vdots \\
\hat{r}(L)
\end{bmatrix}
\]

where

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

- Approximate the observed data sequence \{x[0], ..., x[N]\} with an AR model (consider real-valued process here for simplicity)
- Use biased ACF estimate here to ensure nonnegative definiteness and smaller variance than unbiased estimate (dividing by N-k)
MA Spectral Estimation

An MA(q) model

\[ x[n] = \sum_{k=0}^{q} b_k v[n - k] \quad \Rightarrow \quad B(z) = \sum_{k=0}^{q} b_k z^{-k} \]

can be used to define an MA spectral estimator

\[ \hat{P}_{MA}(f) = \sigma^2 \left| 1 + \sum_{k=1}^{q} b_k e^{-j2\pi fk} \right|^2 \]

Recall:

(1) The problem of solving for \( b_k \) given \( \{r(k)\} \) is to solve a set of nonlinear equations;

(2) An MA process can be approximated by an AR process of sufficiently high order.
Basic Idea to Avoid Solving Nonlinear Equations

Consider two processes:

- **Process#1:** we observed $N$ samples, and need to perform spectral estimate
  - We first model it as a high-order AR process, generated by $1/A(z)$ filter

- **Process#2:** an MA-process generated by $A(z)$ filter
  - Since we know $A(z)$, we can know process#2’s autocorrelation function;
  - We model process#2 as an AR(q) process => the filter would be $1/B(z)$
Complex Exponentials in Additive Noise

$$X[n] = \sum_{i=1}^{p} A_i e^{j\phi_i} e^{j2\pi f_i n} + W[n]$$

$$n = 0, 1, \ldots, N-1 \text{ (Observe N samples)}$$

$$W[n] \text{ — white noise, zero mean, variance } \sigma_w^2$$

$$A_i, f_i \text{ — real, constant, unknown}$$

$$\Rightarrow \text{ to be estimated}$$

$$\phi_i \text{ — uniform distribution over } [0, 2\pi); \text{ uncorrelated with } W[n] \text{ and between different } i.$$
Correlation Matrix for the Process

– Determine autocorrelation function

– Rs = ?  Rw = ?  Rx = ?

– Rank of correlation matrices?
Correlation Matrix for the Process

\[
\Gamma_x(K) = E[x_cn^H x_cn-K] = \sum_{i=1}^{P} A_i \sum_{m=1}^{\Delta} e^{j2\pi f_i k} + \sigma_w^2 \delta(k) = \Pi_i
\]

An M x M correlation matrix for \{x_cn\} (M > P):

\[
R_x = R_s + R_w
\]

\[
R_w = \sigma_w^2 I \quad \text{full rank}
\]

\[
R_s = \sum_{i=1}^{P} \Pi_i e_i e_i^H
\]

where \(e_i = [1, e^{j2\pi f_i}, e^{j4\pi f_i}, \ldots, e^{j2\pi f_i(M-1)}]^T\)
Correlation Matrix for the Process

\[ x[n] = A \exp [j(2\pi f_0 n + \phi)] \]
\[ \mathbb{E}[x[n]] = 0 \quad \forall n \]
\[ \mathbb{E}[x[n] x[n-K]] = \mathbb{E}[A \exp [j(2\pi f_0 n + \phi)] \cdot A \exp [j(2\pi f_0 n - 2\pi f_0 K + \phi)]] \]
\[ = A^2 \exp [j(2\pi f_0 K)] \]
\[ \therefore \text{ } x[n] \text{ is zero-mean w.s.s. with } \Gamma_x(k) = A^2 \exp (j2\pi f_0 K). \]

\[ y[n] = x[n] + w[n] \quad \text{white noise: } \mathbb{E}[w[n] w^*[n-K]] = \begin{cases} \sigma^2 & k = 0 \\ 0 & \text{else} \end{cases} \]
\[ \Gamma_y(k) = \mathbb{E}[y[n] y^*[n-K]] = \mathbb{E}[(x[n] + w[n])(x^*[n-K] + w^*[n-K])] \]
\[ = \Gamma_x(k) + \Gamma_w(k) \quad (\because \mathbb{E}[x[n] w[n]] = 0 \text{ uncorrelated}) \]
\[ = A^2 \exp (j2\pi f_0 K) + \sigma^2 S(k) \]

\[ \mathbb{E}[x(\cdot) w(\cdot)] = \mathbb{E}[x(\cdot)] \mathbb{E}[w(\cdot)] = 0 \]
\[ \text{this crosscorr term vanish because of uncorrelated *and* zero mean for either } x(\cdot) \text{ or } w(\cdot). \]