Statistical Signal Processing

8. Parametric Methods for Spectral 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 Related Readings on Part-III

Overview   Haykins 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

Motivation

- Implicit assumption by classical methods
  - Classical methods use Fourier transform on either *windowed* data/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
  - We 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 us 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 obtain estimates $\hat{r}(k)$ for small k's
  - $\hat{r}(k)$ of larger lags are implicitly extrapolated by the estimated model
- Relation between $r(k)$ and filter parameters $\{a_k\}$ and $\{b_k\}$
  - Related by Yale-Walker equations
  - Solve the equations using $\hat{r}(k)$ to obtain $\{\hat{a}_k\}$ and $\{\hat{b}_k\}$
  - Plug $\{\hat{a}_k\}$ and $\{\hat{b}_k\}$ into $H(z)$ to obtain the estimated PSD, $\hat{P}(\omega)$.
- Deal with MA's nonlinear parameter equations
  - Try to convert/relate them to the AR models that have linear equations

Review: Parameter Equations

Yule-Walker equations (for AR process)

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

$$\begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-p+1) \\ r_x(1) & r_x(0) & \cdots & r_x(-p+2) \\ \vdots & & \ddots & \vdots \\ r(p-1) & \cdots & & r(0) \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} = - \begin{bmatrix} r_x(1) \\ r_x(2) \\ \vdots \\ r_x(p) \end{bmatrix}$$

ARMA model

$$r_x[k] = \begin{cases} -\sum_{l=1}^{p} a[l] r_x[k-l] + \sigma^2 \sum_{l=0}^{q-k} h^*[l] b[l+k] & k=0,1,\cdots q \\ -\sum_{l=1}^{p} a[l] r_x[k-l], & k \geq q+1. \end{cases}$$

MA model

$$r_x[k] = \begin{cases} \sigma^2 \sum_{l=0}^{q-k} b^*[l] b[l+k] & \text{for } k=0,1,\cdots q \\ 0 & \text{for } k \geq q+1 \end{cases}$$

8.1 AR Spectral Estimation

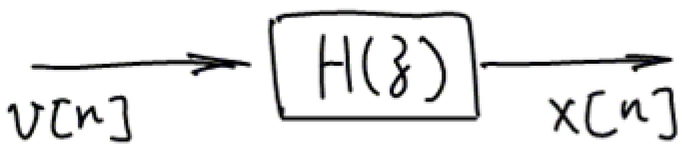
(1) Review of AR process

– The time series $\{x[n], x[n-1], \ldots, x[n-m]\}$ is a realization of an AR process of order M if it satisfies difference equation

$$x[n] + a_1 x[n-1] + \ldots + a_M x[n-M] = v[n]$$

where $\{v[n]\}$ is a white noise process with variance σ^2.

– Generating an AR process with parameters $\{\hat{a}_i\}$:



$$\hat{H}(z) = \frac{1}{1 + \sum_{i=1}^{M} \hat{a}_i z^{-i}} = \frac{1}{\hat{A}(z)}$$

P.S.D. of AR Process

The estimated PSD of an AR process $\{x[n]\}$ is given by

$$\hat{P}_{\mathrm{AR}}(z) = \frac{\sigma^2}{\hat{A}(z)\hat{A}^*(1/z^*)}$$

$$\Downarrow \quad z = e^{j\omega} = e^{j2\pi f}$$

$$\hat{P}_{\mathrm{AR}}(f) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^{M} \hat{a}_k e^{-j2\pi f k}\right|^2}$$

Procedure of AR Spectral Estimation

- Observe the available data points $x[0], \ldots, x[N\text{-}1]$, and Determine the AR process order p
- Estimate the autocorrelation functions (ACF) $k = 0, \ldots, p$

Biased (low variance)

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

Unbiased (may not non-neg. definite)

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

- Solve $\{\hat{a}_i\}$ from the Yule-Walker equations (or the normal equations of forward linear prediction)
  - Recall for an AR process, the normal equation of FLP is equivalent to the Yule-Walker equation
- Obtain estimated power spectrum:

$$\hat{P}_{\text{AR}}(f) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_k e^{-j2\pi fk}\right|^2}$$

8.2 Maximum Entropy Spectral Estimation (MESE)

- Viewpoint: Extrapolations of ACF
  - $\{\hat{r}[0], \ldots \hat{r}[p]\}$ is known; there are generally an infinite number of possible extrapolations for $r(k)$ at larger lags
  - As long as $\{r[p+1], r[p+2], \ldots\}$ guarantee that the correlation matrix is non-negative definite, they all form valid ACFs for w.s.s.
- Maximum entropy principle
  - Perform extrapolation s.t. the time series (characterized by the extrapolated ACF) has maximum entropy
  - i.e., the time series will be the least constrained thus most random one among all series having the same first $(p+1)$ ACF values

=> Maximizing entropy leads to estimated PSD be the smoothest one

- Recall white noise process has flat PSD

MESE for Gaussian Process: Formulation

For a Gaussian random process, the entropy per sample is proportional to

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P(f)df$$

Thus the max entropy spectral estimation is

$$\max \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P(f)df$$

subject to

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} P(f)e^{j2\pi fk}df = \hat{r}(k), \qquad \text{for } k = 0,1,\ldots,p$$

MESE for Gaussian Process: Solution

Using the Lagrangian multiplier technique, the solution can be found as

$$\hat{P}_{\mathrm{ME}}(f) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_k e^{-j2\pi f k}\right|^2}$$

where $\{\hat{a}_k\}$ are found by solving the Yule-Walker equations given the estimated ACF values $\hat{r}[0], \ldots \hat{r}[p]$.

- For Gaussian processes, the MESE is equivalent to AR spectral estimator and the $\hat{P}_{\mathrm{ME}}(f)$ is an all-pole spectrum
  - Different assumptions on the process: Gaussian vs. AR processes

8.3 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

$$\hat{P}_{\mathrm{MA}}(f) = \sigma^2 \left| 1 + \sum_{k=1}^{q} \hat{b}_k e^{-j2\pi f k} \right|^2$$

Recall important results on MA process:

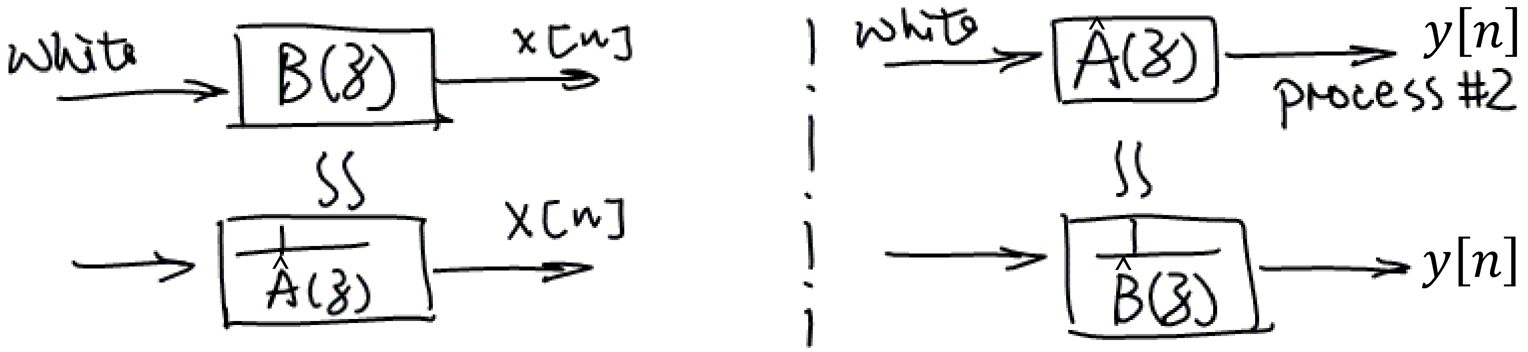
(1) The problem of solving for b_k given $\{r(k)\}$ is to solve a set of nonlinear equations;

(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: an approximated high-order AR process in the observed data $x[n]$
  - We model $x[n]$ as a high-order AR process generated by $1/\hat{A}(z)$ filter



- Process #2: an MA process $y[n]$ generated by $\hat{A}(z)$ filter
  - Since we know $\hat{A}(z)$, we can obtain $y[n]$'s autocorrelation values $r_y(k)$
  - We model process #2 as an AR(q) process => the filter would be $1/\hat{B}(z)$

Use AR Model to Help Finding MA Parameters

– For simplicity, we consider the real coefficients for the MA model.

Note $$P_{MA}(z) = \sigma^2 B(z) B(z^{-1})$$

To approximate it with an AR(L) model, i.e.,

$$P_{MA}(z) \approx \frac{\sigma^2}{\hat{A}(z)\hat{A}(z^{-1})}$$

where $\hat{A}(z) = 1 + \sum_{k=1}^{L} \hat{a}_k z^{-k}$

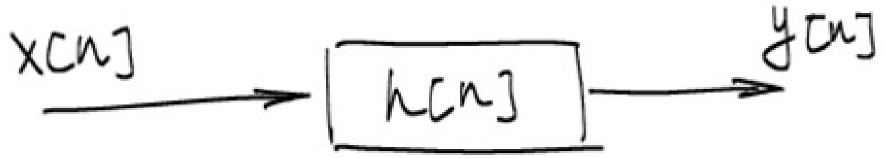
$L >> q$

$$\Rightarrow \underbrace{\hat{A}(z)\hat{A}(z^{-1})}_{\text{order } L} \approx \frac{1}{\underbrace{B(z)B(z^{-1})}_{\text{order } q}}$$

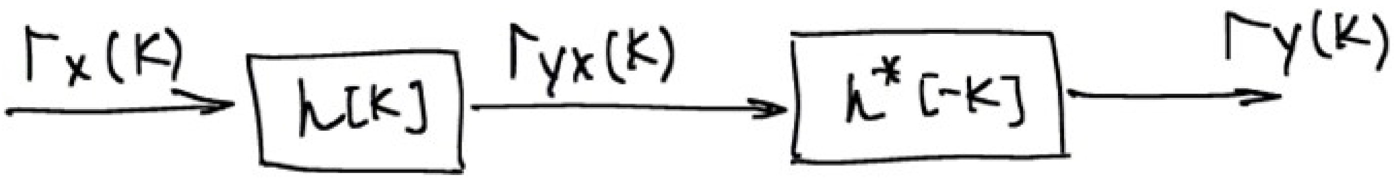
- The RHS represents power spectrum of an AR(q) process
- The inverse ZT of LHS is the ACF of the AR(q) process

Recall: ACF of Output Process After LTI Filtering

w.s.s. process



stable LTI filter



filter filter

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

$$\Downarrow \text{ZT}$$

$$H(z)\, H^*(1/z^*)$$

Use AR to Help Finding MA Parameters (cont'd)

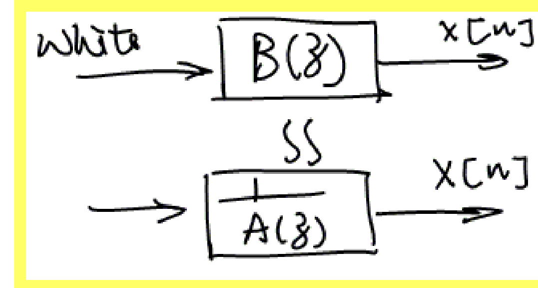
Let $x[n] = w[n] \sim N(0, \sigma_w^2)$ i.i.d., and $h[n] = \hat{a}_n$, we have

$$r_y(k) = \sigma_w^2 \sum_{n=0}^{L-k} \hat{a}_n \hat{a}_{n+k} \qquad \text{for lag } k$$

➔ Knowing autocorrelation sequence $r_y(k)$, the best AR coefficients $\{\hat{b}_k\}$ for process #2 can be obtained by direct matrix inverse or Levinson-Durbin recursion.

- Note that the best AR coefficients for process #2 are actually the best MA coefficients for process #1.

Durbin's Method



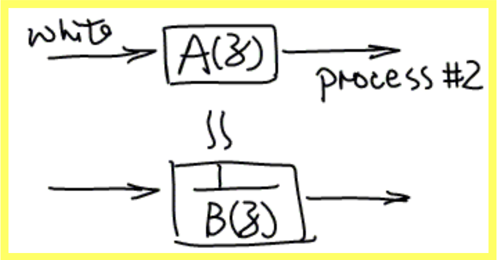
1. Use Levinson-Durbin recursion and solve for

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

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

- We first approximate the observed data sequence $\{x[0], \ldots, x[N]\}$ with an AR model of high order (often pick $L > 4q$)
- We use biased ACF estimator ($1/N$) to ensure nonnegative definiteness and smaller variance than unbiased estimator $[1/(N-k)]$

Durbin Method (cont'd)



2. Fit an AR(q) model to the data sequence $\{1, \hat{a}_1, \hat{a}_2, \ldots, \hat{a}_L\}$

$$\begin{bmatrix} \hat{r}_a(0) & \hat{r}_a(1) & \cdots & \hat{r}_a(q-1) \\ \hat{r}_a(1) & \hat{r}_a(0) & & \hat{r}_a(q-2) \\ \vdots & & \ddots & \vdots \\ \hat{r}_a(q-1) & \cdots & \cdots & \hat{r}_a(0) \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_q \end{bmatrix} = - \begin{bmatrix} \hat{r}_a(1) \\ \vdots \\ \hat{r}_a(q) \end{bmatrix}$$

where $$\hat{r}_a(K) = \frac{1}{L+1} \sum_{n=0}^{L-K} \hat{a}_n \hat{a}_{n+K}$$

– The result $\{b_i\}$ is the estimated MA parameters for original $\{x[n]\}$

– Note we add $1/(L+1)$ factor to allow the interpretation of $r_a(k)$ as an autocorrelation function estimator

8.4 ARMA Spectral Estimation

Recall the ARMA(p, q) model

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + \sum_{k=0}^{q} b_k v[n-k]$$

We define an ARMA(p, q) spectral estimator

$$\hat{P}_{\mathrm{ARMA}}(f) = \hat{\sigma}^2 \frac{\left|1 + \sum_{k=1}^{q} \hat{b}_k e^{-j2\pi fk}\right|^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_k e^{-j2\pi fk}\right|^2}$$

Modified Yule-Walker Equations

Recall the Yule-Walker Eqs. for ARMA(p, q) process

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

We may use equations for $k \geq q+1$ to solve for $\{a_l\}$

$$
\begin{bmatrix}
r(q) & r(q-1) & \cdots & r(q-p+1) \\
r(q+1) & r(q) & & \vdots \\
\vdots & & \ddots & \\
r(q+p-1) & \cdots & \cdots & r(q)
\end{bmatrix}
\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}
= -
\begin{bmatrix} r(q+1) \\ \vdots \\ r(q+p) \end{bmatrix}
$$

$\Rightarrow S \underline{\hat{a}} = \underline{t}$ "Modified Yule-Walker Equations"

Estimating ARMA Parameters

1. By solving the modified Yule-Walker eqs., we obtain

$$\hat{A}(z) = 1 + \sum_{k=1}^{p} \hat{a}_k z^{-k}$$

2. We eliminate the AR component by filtering $x[n]$ with FIR filter $\hat{A}(z)$ to obtain an approximate MA(q) process:

$$\hat{A}(z)X(z) = \hat{A}(z)\frac{B(z)}{\mathrm{A}(z)}W(z) \approx B(z)W(z)$$

3. Coefficients $\{b_k\}$ can be estimated by Durbin's method.

Extension: LSMYWE Estimator

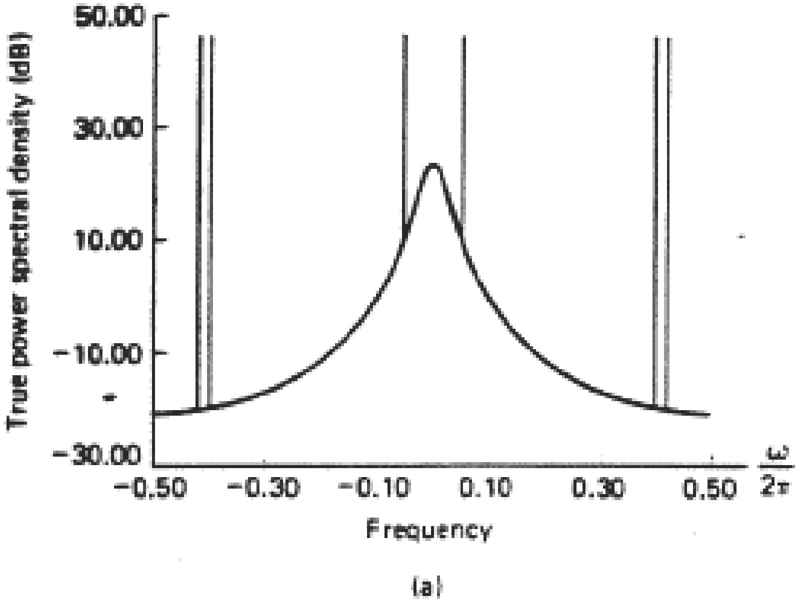
- Performance by solving *p* modified Yule-Walker equations followed by Durbin's method
  - May yield highly noisy spectral estimates (esp. when the matrix involving ACF is nearly singular due to poor ACF estimates)
- Improvement: use more than *p* equations to solve $\{\hat{a}_1, \ldots, \hat{a}_p\}$ in a least squared sense
  - Use Yule-Walker equations for $k = (q+1), \ldots, M$: $\min \|\mathbf{t} - \mathbf{S}\mathbf{a}\|^2$
  - Least-squares solution: $\hat{\mathbf{a}} = (\mathbf{S}^H\mathbf{S})^{-1}\mathbf{S}^H\mathbf{t}$
  - Then obtain $\{b_k\}$ by Durbin's method

➔ "Least-Squares Modified Yule-Walker Equations" (LSMYWE)

Ref: review in Hayes' book Sec.2.3.6 on least square solution

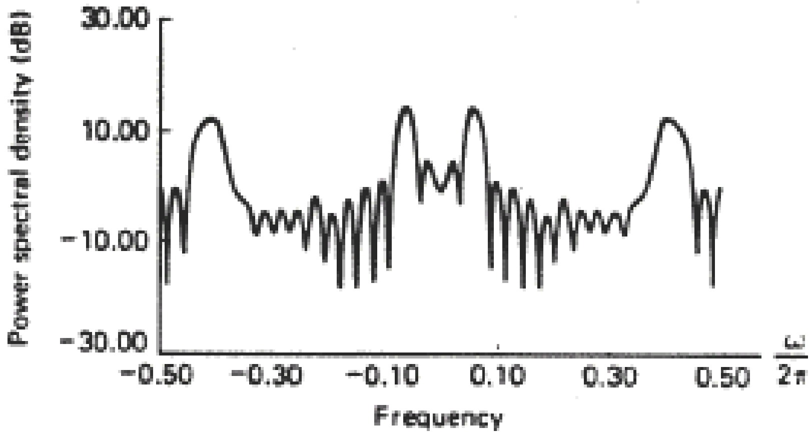
Comparison of Different Methods: Revisit

- 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 $f = 1/32$



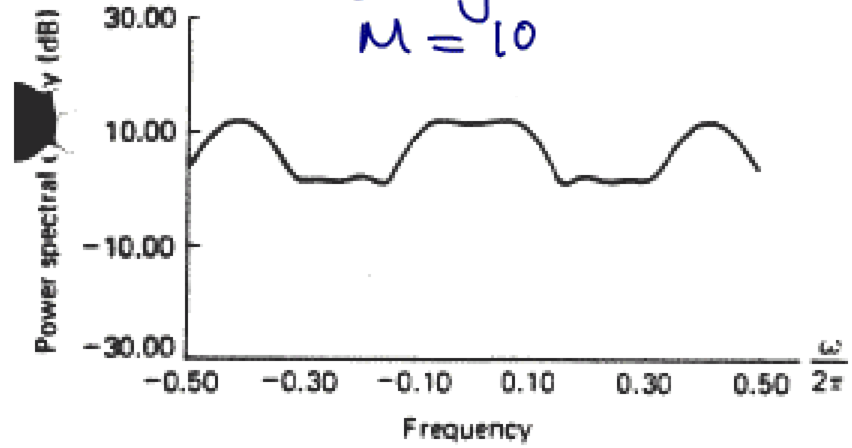
- Examine typical characteristics of various non-parametric and parametric spectral estimators

(Fig.2.17 from Lim/Oppenheim book)

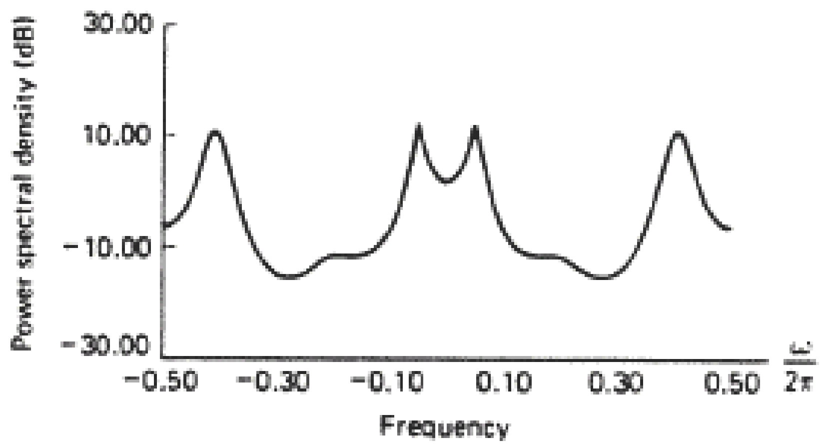


(b) Periodogram

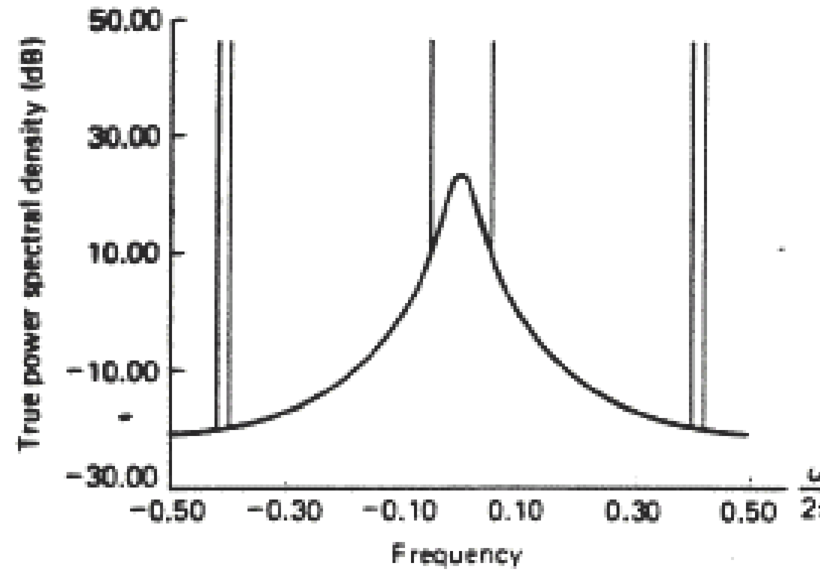
triangle window
M = 10



(c) Blackman-Tukey

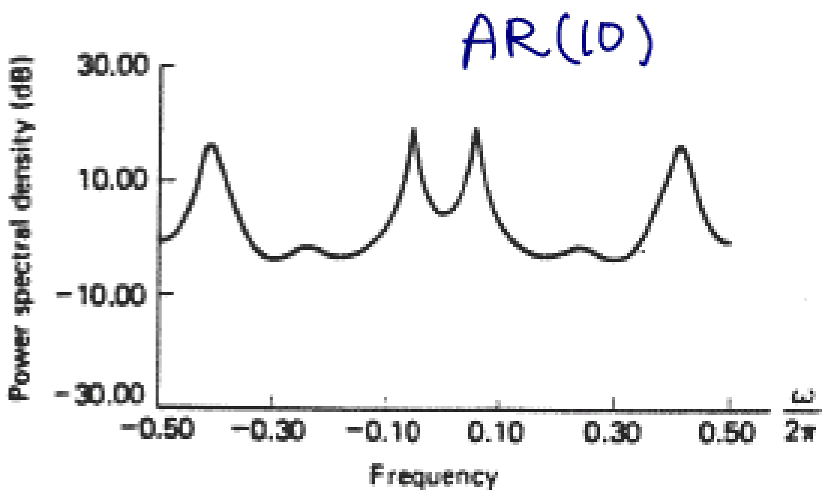


(d) Minimum variance spectral estimator

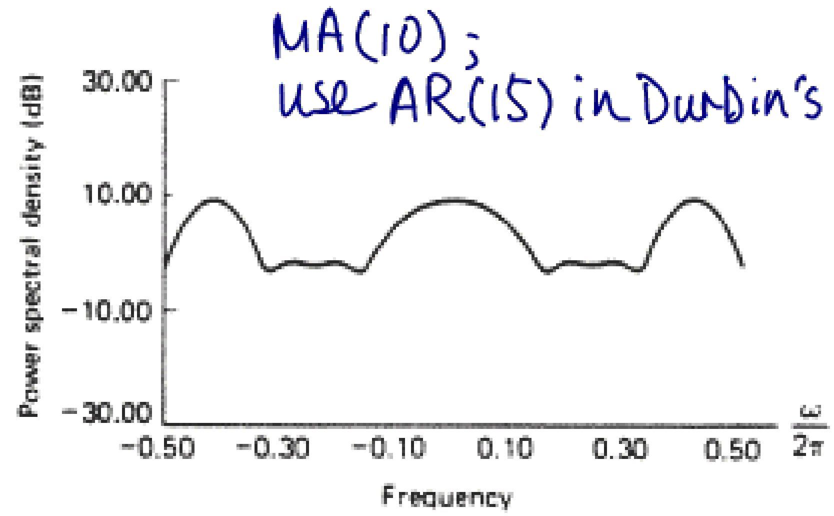


true p.s.d.

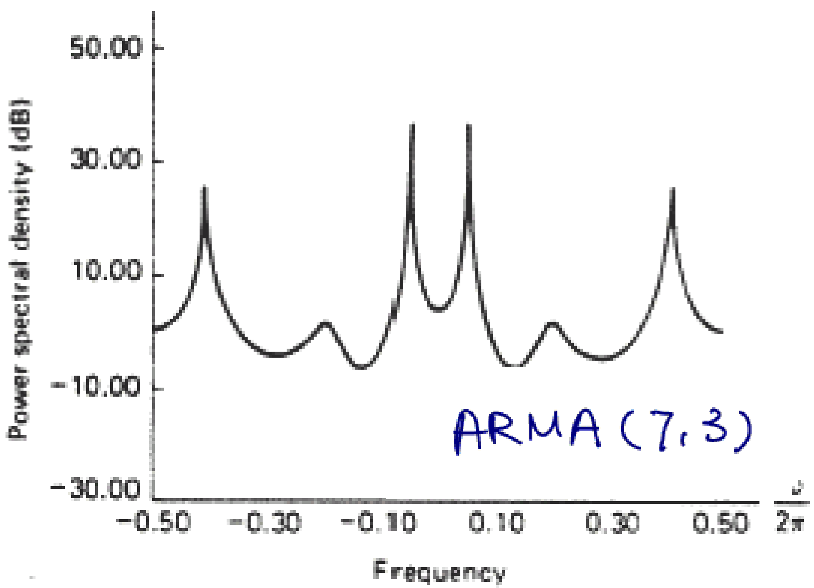
(a)



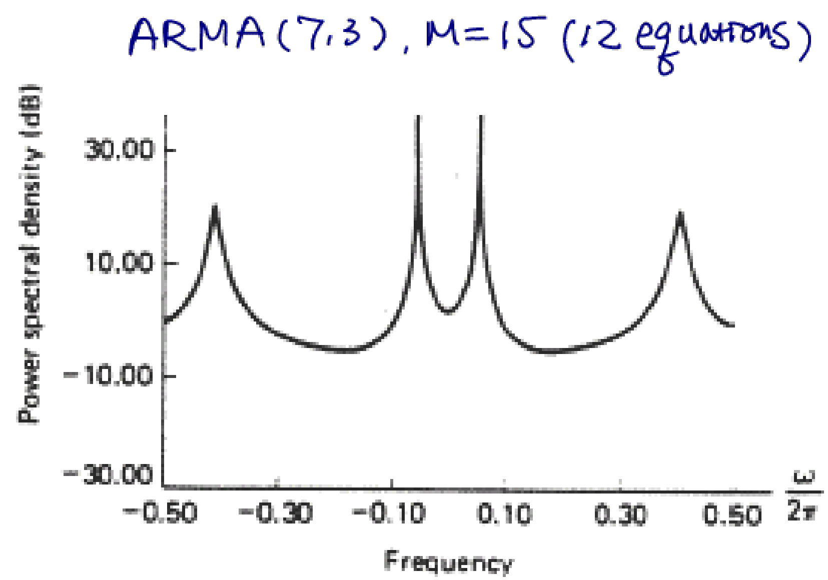
(e) Autocorrelation



(i) Durbin



(j) Modified Yule-Walker equations



(k) Least-squares modified Yule-Walker equations

8.5 Model Order Selection

- The best way to determine the model order is to base it on the physics of the data generation process
- Example: speech processing
  - Studies show the vocal tract can be modeled as an all-pole filter having 4 resonances in a 4kHz band, thus at least 4 pairs of complex conjugate poles are necessary

    ➔ Typically 10–12 poles are used in an AR modeling for speech
- When no such knowledge is available, we can use some statistical test to estimate the order

Ref. for in-depth exploration: "Model-order selection," by P. Stoica and Y. Selen, IEEE Signal Processing Magazine, July 2004.

Considerations for Order Selection

- Modeling error
  - Modeling error measures the (statistical) difference between the true data value and the approximation by the model

    e.g., estimating linear prediction MSE in AR modeling
  - Usually for a given type of model (e.g., AR, ARMA), the modeling error decreases as we increase the model order
- Balance between the modeling error and the amount of model parameters to be estimated
  - The number of parameters that need to be estimated and represented increases as we use higher model order ➔ Cost of overmodeling
  - Can balance modeling error and the cost of going to higher model by imposing a penalty term that increases with the model order

A Few Commonly Used Criteria

- Akaike Information Criterion (AIC)
  - A general estimate of the Kullback-Leibler divergence between assumed and true p.d.f., with an order penalty term increasing linearly
  - Choose the model order that minimize AIC

$$\mathrm{AIC}(i) = N \ln \varepsilon_p + 2i$$

size of available data (N)

model error (ε_p)

model order: (i)
i=*p* for AR(*p*)
i=*p*+*q* for ARMA(*p*, *q*)

- Minimum Description Length (MDL) Criterion
  - Impose a bigger penalty term to overcome AIC's overestimation
  - Estimated order converges to the true order as N goes to infinity

$$\mathrm{MDL}(i) = N \ln \varepsilon_p + (\log N) i$$