Part 3. Spectrum Estimation

3.2 Parametric Methods for Spectral 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*

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

Motivation

- 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
  - 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 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)
  - 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/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

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

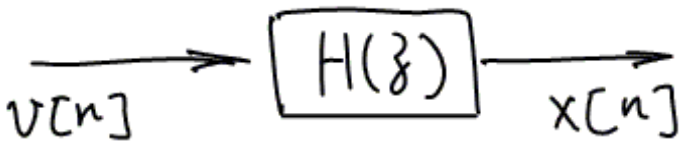
3.2.1 AR Spectral Estimation

(1) Review of AR process

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

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

– Generating an AR process with parameters $\{a_i\}$:



3.2.1 AR Spectral Estimation

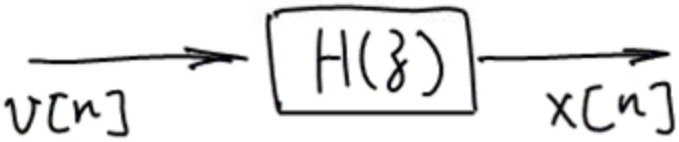
(1) Review of AR process

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



$$H(z) = \frac{1}{1 + \sum_{i=1}^{M} a_i z^{-i}} \overset{\text{def}}{=} \frac{1}{A(z)}$$

P.S.D. of An AR Process

Recall: the p.s.d. of an AR process {x[n]} is given by

$$\hat{P}_{\mathrm{AR}}(z) =$$

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

$$\hat{P}_{\mathrm{AR}}(f) =$$

P.S.D. of An AR Process

Recall: the p.s.d. of an AR process {x[n]} is given by

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

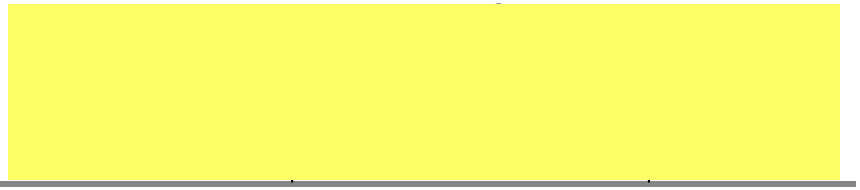
$$\Downarrow \quad z = e^{j\omega} = e^{j2\pi f}$$

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

Procedure of AR Spectral Estimation

- Observe the available data points x[0], …, x[N-1], and Determine the AR process order *p*
- Estimate the autocorrelation functions (ACF) k=0,…p

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



Procedure of AR Spectral Estimation

- Observe the available data points x[0], …, x[N-1], and Determine the AR process order *p*
- Estimate the autocorrelation functions (ACF) k=0,…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 { a_i } from the Yule-Walker equations or the normal equation of forward linear prediction
  - Recall for an AR process, the normal equation of FLP is equivalent to the Yule-Walker equation
- Obtain power spectrum $P_{AR}(f)$:

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

3.2.2 Maximum Entropy Spectral Estimation (MESE)

- View point: Extrapolations of ACF
  - {r[0], r[1], …, 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], … } 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

-

3.2.2 Maximum Entropy Spectral Estimation (MESE)

- Extrapolations of ACF
  - {r[0], r[1], …, 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], … } 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 p.s.d. be the smoothest one

  - Recall white noise process has flat p.s.d.

MESE for Gaussian Process: Formulation

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

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

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

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

MESE for Gaussian Process: Solution

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

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

MESE for Gaussian Process: Solution

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

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

where $\{a_k\}$ are found by solving the Yule-Walker equations given the ACF values r(0), …, r(p)

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

3.2.3 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

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

Recall important results on MA process:

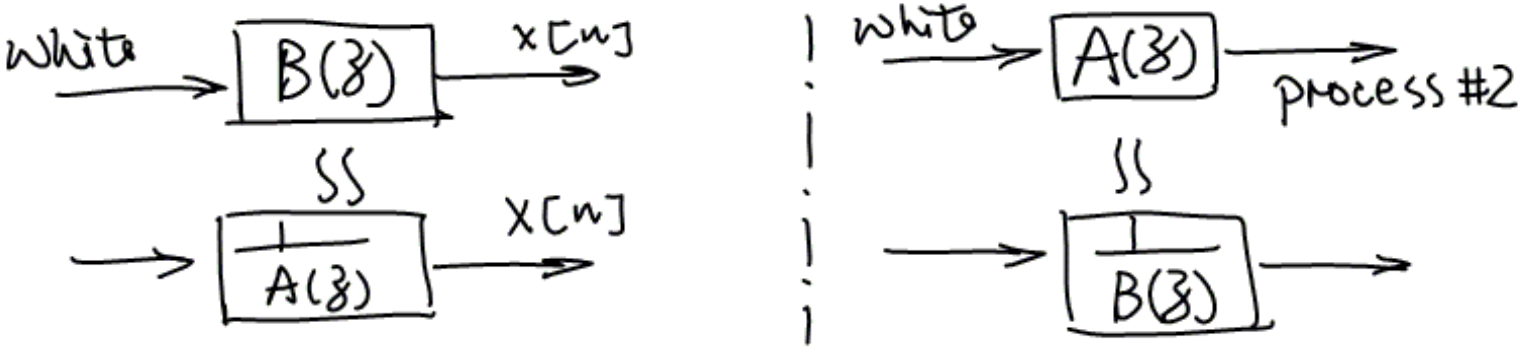
(1) The problem of solving for b_k given {r(k)} is to solve a set of nonlinear equations;

(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: 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)

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.,

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

$$L >> q$$

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.,

$$P_{MA}(z) \approx \frac{\sigma^2}{A(z)A(z^{-1})} \qquad \text{where } A(z) = 1 + \sum_{k=1}^{L} a_k z^{-k}, \quad L >> q$$

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

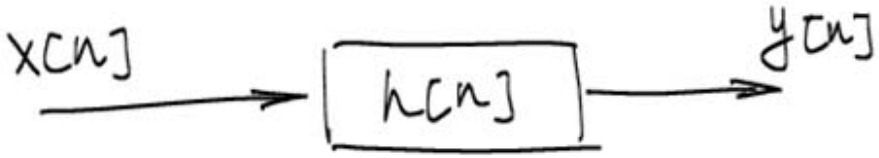
- The RHS represents power spectrum of an AR(q) process
- The unverse ZT of LHS is the ACF of the AR(q) process

Use AR Model to Find MA Parameters: Solutions

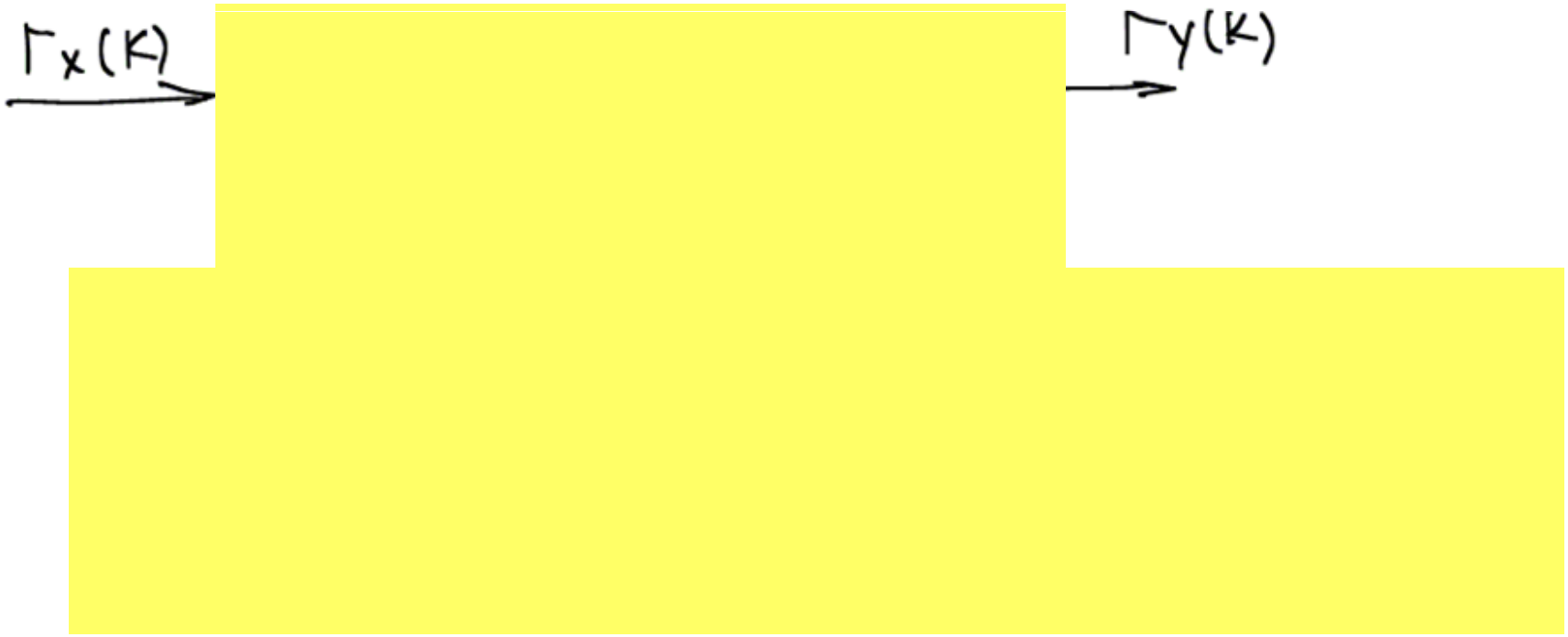
- For simplicity, we consider the real coefficients for the MA model.

Recall: ACF of Output Process After LTI Filtering

w.s.s. process

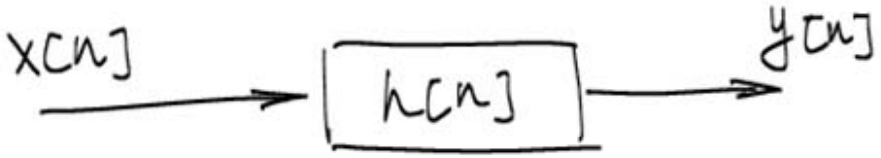


stable LTI filter

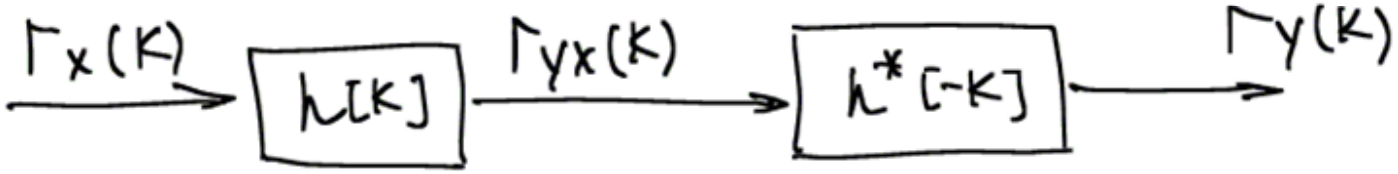


Recall: ACF of Output Process After LTI Filtering

w.s.s. process



stable LTI filter



filter filter

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

$\Downarrow ZT$

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

deterministic autocorrelation of filter's impulse response

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

A random process with power spectrum $A(z)A(z^{-1})$ can be viewed as filtering a white process by a filter $A(z)$, and has autocorrelation

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

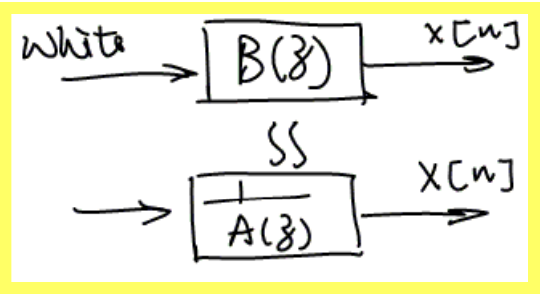
A random process with power spectrum $A(z)A(z^{-1})$ can be viewed as filtering a white process by a filter A(z), and has autocorrelation

proportional to $\sum_{n=0}^{L-k} a_n a_{n+k}$ for lag k

➔ Knowing such autocorrelation function,
we can use Levinson-Durbin recursion to find the optimal linear prediction parameters for the process
(or equivalently its AR approximation parameters)

Thus we get $\{b_k\}$ as $A(z)A(z^{-1}) \approx \frac{1}{B(z)B(z^{-1})}$

Durbin's Method

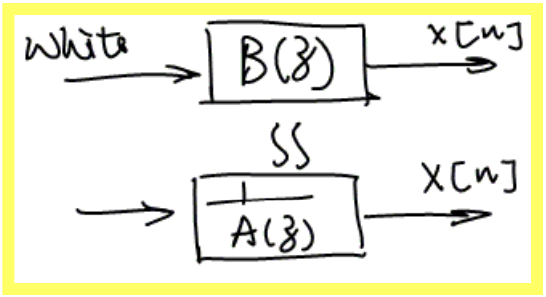


1. Use Levinson-Durbin recursion and solve for

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

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

Durbin's Method



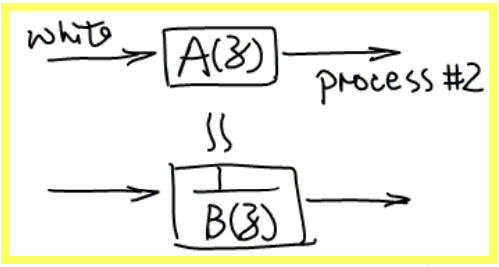
1. Use Levinson-Durbin recursion and solve for

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

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], …, x[N]} with an AR model of high order (often pick L > 4q)
- We use biased ACF estimate here to ensure nonnegative definiteness and smaller variance than unbiased estimate (dividing by N-k)

Durbin Method (cont'd)

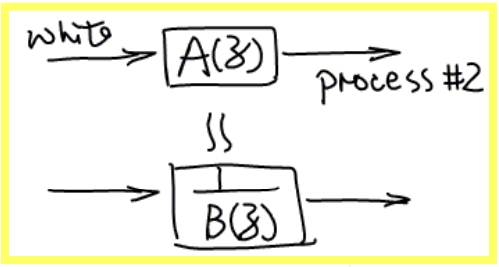


2. Fit the data sequence $\{1, \hat{a}_1, \hat{a}_2, ..., \hat{a}_L\}$ to an AR(q) model:

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

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

Durbin Method (cont'd)



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

$$\begin{bmatrix} \hat{r}_a(0) & \hat{r}_a(1) & \cdots & \hat{r}_a(q-1) \\ \hat{r}_a(1) & \hat{r}_a(0) & & \hat{r}_a(q-2) \\ \vdots & & \ddots & \vdots \\ \hat{r}_a(q-1) & \cdots & & \hat{r}_a(0) \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_q \end{bmatrix} = - \begin{bmatrix} \hat{r}_a(1) \\ \vdots \\ \hat{r}_a(q) \end{bmatrix}$$

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

3.2.4 ARMA Spectral Estimation

Recall the ARMA(p,q) model

We define an ARMA(p,q) spectral estimator

3.2.4 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

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

Modified Yule-Walker Equations

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

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

Modified Yule-Walker Equations

Recall the Yule-Walker Eq. 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 can use the equations for k≥q+1 to solve for $\{a_l\}$

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

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

Estimating ARMA Parameters

1. By solving the modified Yule-Walker eq., we get

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

2. To estimate $\{b_k\}$,

Estimating ARMA Parameters

1. By solving the modified Yule-Walker eq., we get

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

2. To estimate $\{b_k\}$,

   We first filter {x[n]} with $\hat{A}(z)$, and model the output with an MA(q) model using Durbin's method.

Extension: LSMYWE Estimator

- Performance by solving *p* modified Yule-Walker equations followed by Durbin's method
  - May yield highly variable spectral estimates (esp. when the matrix involving ACF is nearly singular due to poor ACF estimates)
- Improvement: use more than p equations to solve $\{a_1 .. a_p\}$ in a least square sense
  - Use Yule-Walker equations for k = (q+1), … M: $\min \| \underline{t} - S \underline{a} \|^2$
  - Least square solution: $\underline{a} = (S^H S)^{-1} S^H \underline{t}$
  - Then obtain $\{b_i\}$ by Durbin's method

➔ "Least-square modified Yule-Walker equation" (LSMYWE)

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

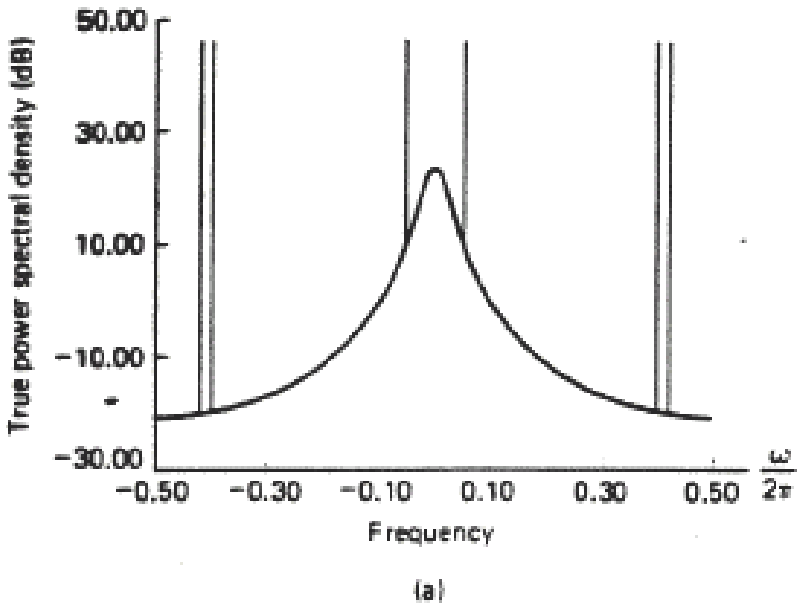
Comparison of Different Methods: Revisit

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

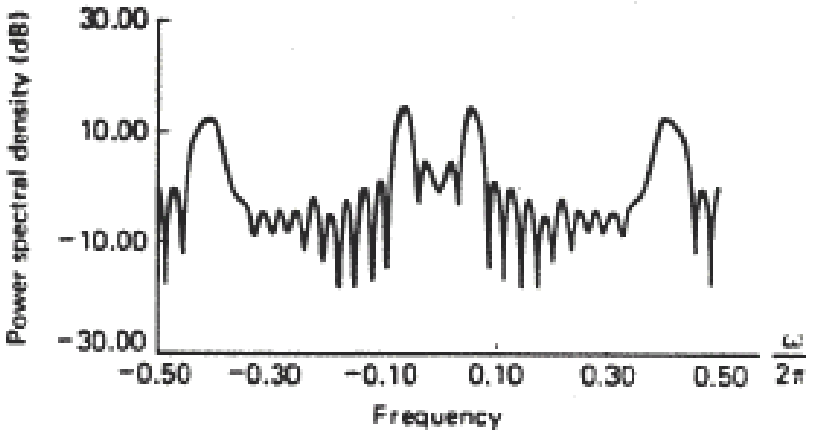
    where $z[n] = -a_1 z[n-1] + v[n]$, $a_1 = -0.85$, $\sigma^2 = 0.1$

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

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

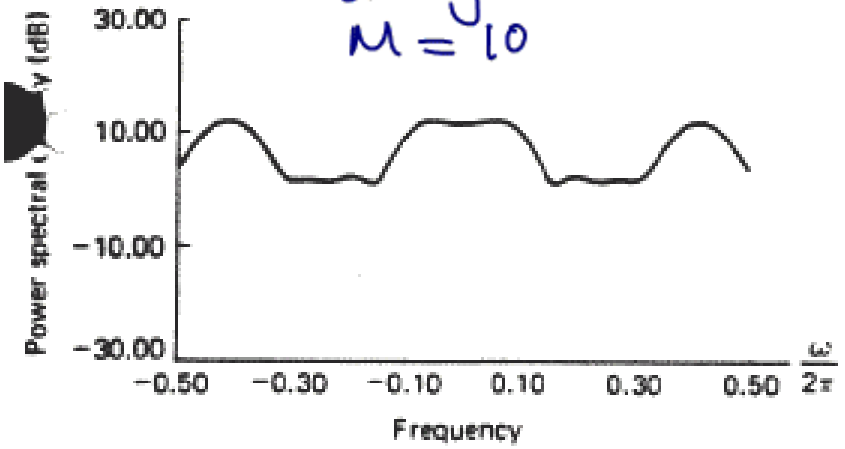


(Fig.2.17 from Lim/Oppenheim book)

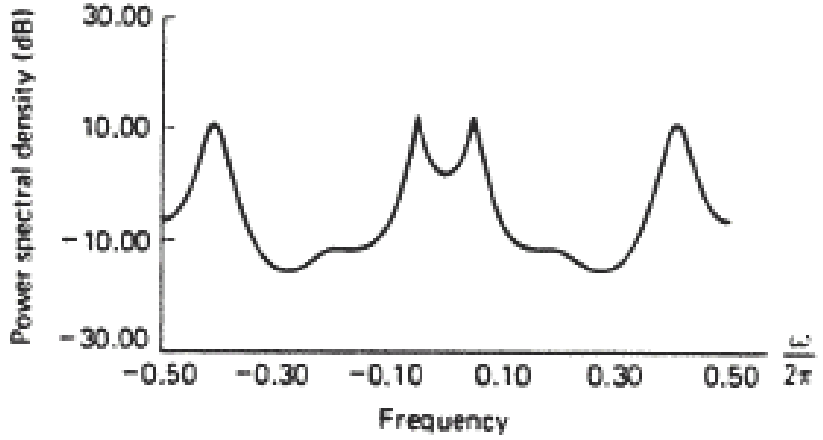


(b) Periodogram

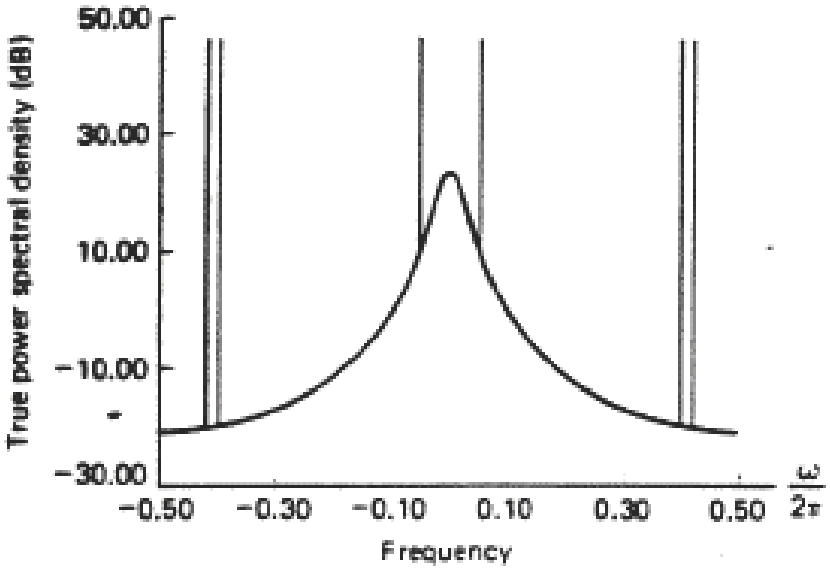
triangle window
M = 10



(c) Blackman-Tukey

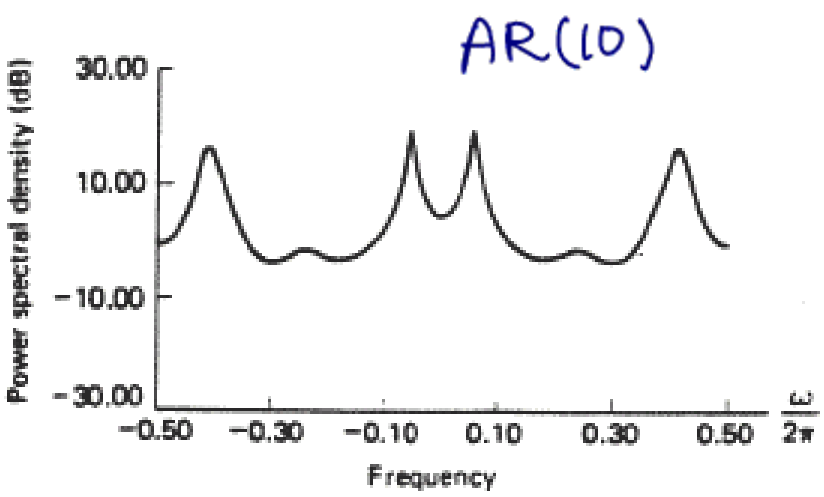


(d) Minimum variance spectral estimator

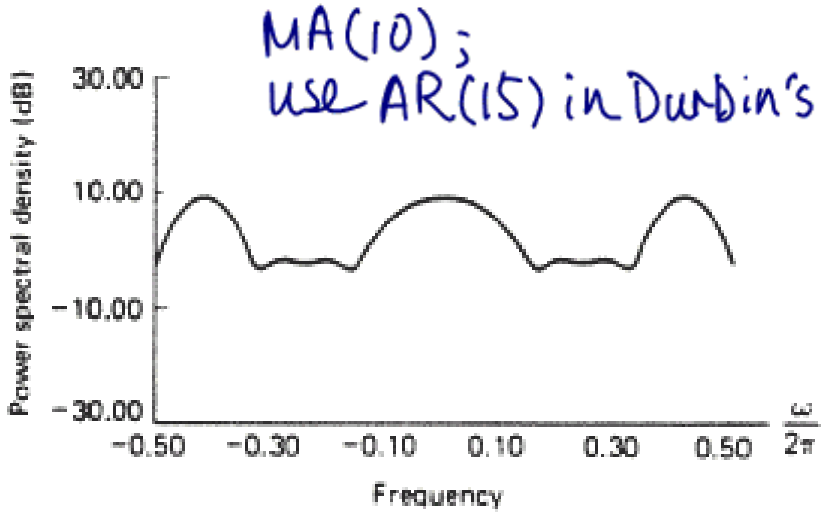


true p.s.d.

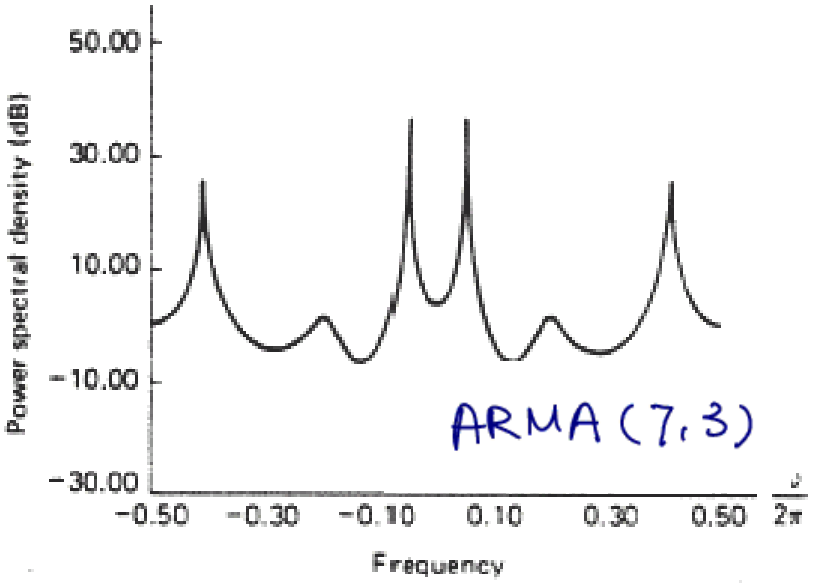
(a)



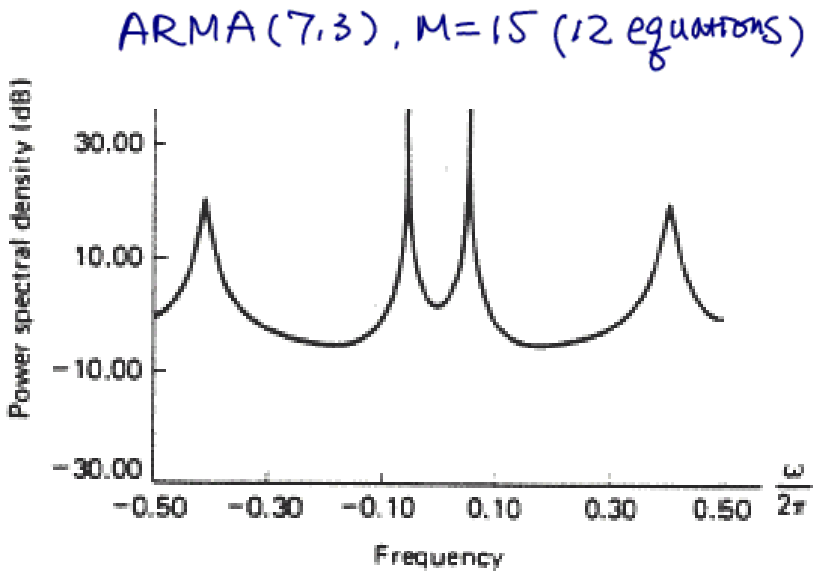
(e) Autocorrelation



(i) Durbin



(j) Modified Yule-Walker equations



(k) Least-squares modified Yule-Walker equations

3.2.5 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

- 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

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

model error

model order:
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

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