ECE7252
Statistical Learning for Signal Processing

Lecture 9: Linear Methods for Regression (Part II: Autoregression)

Chin-Hui Lee
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA
chl@ece.gatech.edu

Lecture Outline on Regression

- Statistical learning: general concept
- Supervised learning: learning with a teacher
- Regression and classification problems
- Model selection, feature selection and generalization
- Autoregression
- Statistical decision theory (in Lecture on classification)

Autoregression

$$x(n) = \sum_{k=1}^{p} \alpha_k \, x(n-k) \text{ for some value of } p, \alpha_k\text{'s}$$

- Autoregressive (AR) model is the most useful technique extensively used in time series analysis and econometrics
- AR (called linear prediction or LP) methods are the most widely used in speech coding, speech synthesis, speech recognition, and speaker recognition
  - LP methods provide extremely accurate estimates of speech parameters, and does it extremely efficiently
  - Basic idea of LP: current speech sample can be closely approximated as a linear combination of past samples.

Basic Principles of LP Speech Analysis

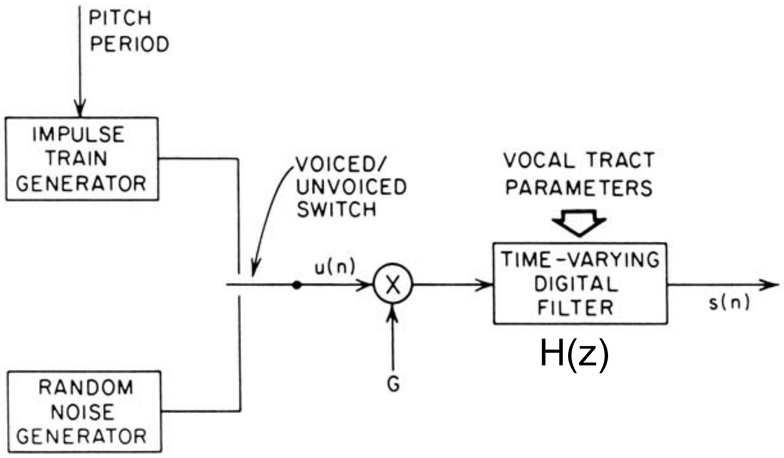


Fig. 8.1 Block diagram of simplified model for speech production.

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + G u(n)$$

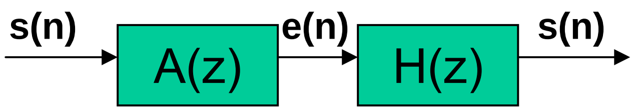
$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

- the time-varying digital filter represents the effects of the glottal pulse shape, the vocal tract IR, and radiation at the lips
- the system is excited by an impulse train for voiced speech, or a random noise sequence for unvoiced speech
- this 'all-pole' model is a natural representation for non-nasal voice speech—but it also works reasonably well for nasals and unvoiced sounds

LP Basic Equations

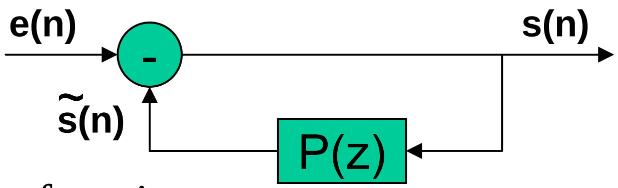
- a p^{th} order linear predictor is a system of the form

$$\tilde{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \Leftrightarrow P(z) = \sum_{k=1}^{p} \alpha_k z^{-k} = \frac{\tilde{S}(z)}{S(z)}$$



- the prediction error, $e(n)$, is of the form

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k)$$



- the prediction error is the output of a system with transfer function

$$A(z) = \frac{E(z)}{S(z)} = 1 - P(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$$

- if the speech signal obeys the production model exactly, and if $\alpha_k = a_k, 1 \le k \le p$

$\Rightarrow e(n) = Gu(n)$ and $\underline{A(z) \text{ is an inverse filter for } H(z)}$, i.e.,

$$H(z) = \frac{1}{A(z)}$$

Solution for $\{\alpha_k\}$

- short-time average prediction error is defined as

$$E_n = \sum_m e_n^2(m) = \sum_m \left(s_n(m) - \tilde{s}_n(m)\right)^2$$

$$= \sum_m \left(s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right)^2$$

- select segment of speech $s_n(m) = s(m+n)$ in the vicinity of sample n
- the key issue to resolve is the range of m for summation (to be discussed later)

Center of Signal and Image Processing
Georgia Institute of Technology

Solution for $\{\alpha_k\}$

- can find values of α_k that minimize E_n by setting:

$$\frac{\partial E_n}{\partial \alpha_i} = 0, \qquad i = 1, 2, \ldots, p$$

- giving the set of equations

$$-2\sum_m s_n(m-i)[s_n(m) - \sum_{k=1}^{p} \hat{\alpha}_k s_n(m-k)] = 0, \quad 1 \le i \le p$$

$$-2\sum_m s_n(m-i)\; e_n(m) = 0, \quad 1 \le i \le p$$

- where $\hat{\alpha}_k$ are the values of α_k that minimize E_n (from now on just use α_k rather than $\hat{\alpha}_k$ for the optimum values)
- prediction error $(e_n(m))$ is orthogonal to signal $(s_n(m-i))$ for delays (i) of 1 to p

Solution for $\{\alpha_k\}$

- defining

$$\phi_n(i,k) = \sum_m s_n(m-i)s_n(m-k)$$

- we get

$$\sum_{k=1}^{p} \alpha_k \phi_n(i,k) = \phi_n(i,0), \quad i = 1,2,\ldots,p$$

- leading to a set of p equations in p unknowns that can be solved in an efficient manner for the $\{\alpha_k\}$

Center of Signal and Image Processing
Georgia Institute of Technology

Autocorrelation Method: Normal Equations

- since $R_n(k)$ is even, then

$$\phi_n(i,k) = R_n(|i-k|),\ 1 \le i \le p,\ 0 \le k \le p$$

- thus the basic equation becomes

$$\sum_{k=1}^{p} \alpha_k \phi_n(i-k) = \phi_n(i,0), \quad 1 \le i \le p$$

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i), \quad 1 \le i \le p$$

- with the minimum mean-squared prediction error of the form

$$E_n = \phi_n(0,0) - \sum_{k=1}^{p} \alpha_k \phi_n(0,k)$$

$$= R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k)$$

Autocorrelation Method: Matrix Equation

- as expressed in matrix form

$$\begin{bmatrix} R_n(0) & R_n(1) & . & . & R_n(p-1) \\ R_n(1) & R_n(0) & . & . & R_n(p-2) \\ . & . & . & . & . \\ . & . & . & . & . \\ R_n(p-1) & R_n(p-2) & . & . & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ . \\ . \\ R_n(p) \end{bmatrix}$$

$$\Re \boldsymbol{\alpha} = \boldsymbol{r}$$

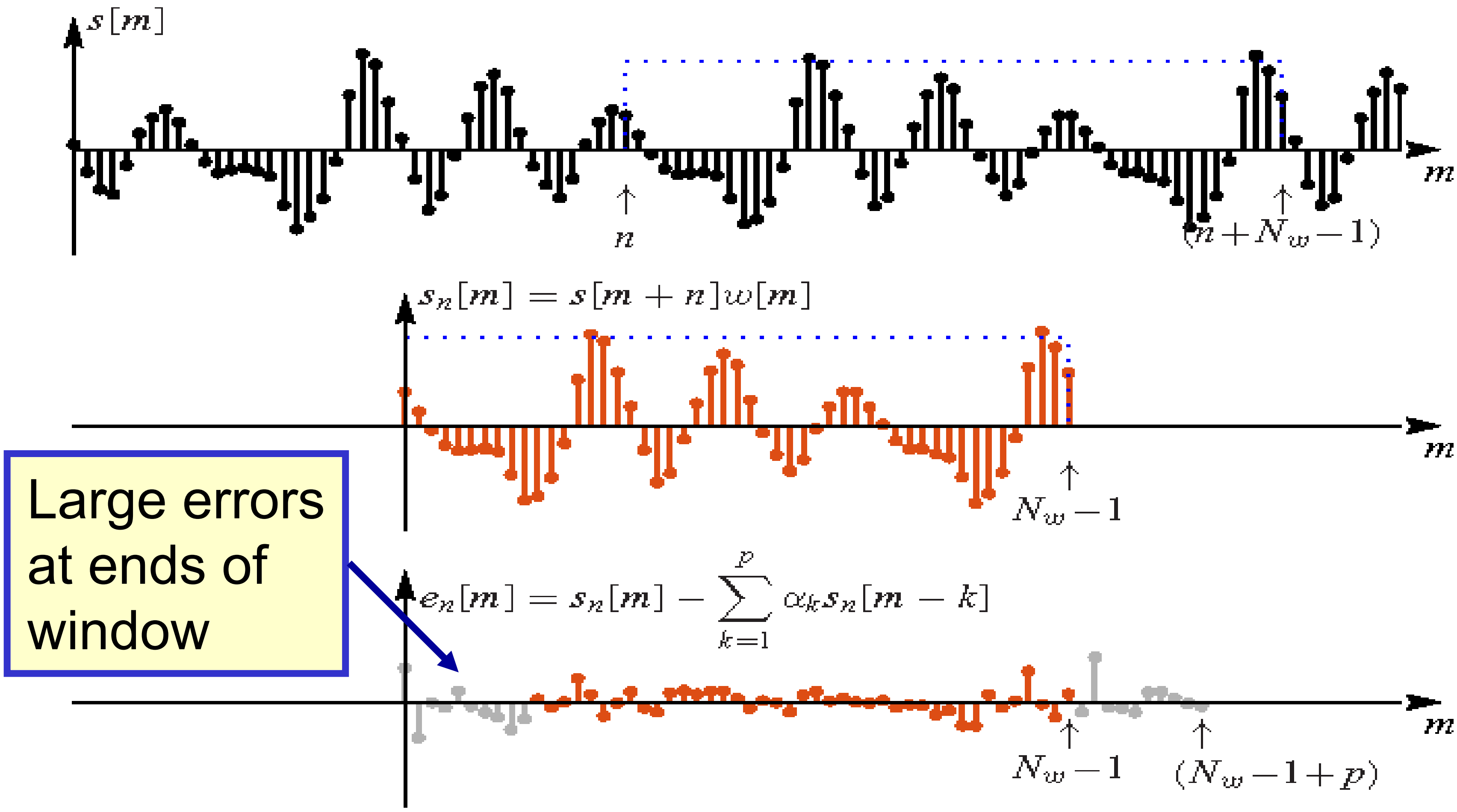
with solution

$$\boldsymbol{\alpha} = \Re^{-1} \boldsymbol{r}$$

- $\Re$ is a $p \times p$ Toeplitz Matrix => symmetric with all diagonal elements equal => there exist more efficient algorithms to solve for $\{\alpha_k\}$ than simple matrix inversion

Center of Signal and Image Processing
Georgia Institute of Technology

Autocorrelation Method with Windows



Solution to Autocorrelation Method

- the matrix equation for the autocorrelation method is

$$\sum_{k=1}^{p} \alpha_k R_n(|i-k|) = R_n(i) \qquad 1 \le i \le p$$

- since matrix is Toeplitz (symmetric with all elements along each diagonal equal) there are several solution methods, including the Levinson method, but we will use the Durbin recursive procedure which goes as follows:

$$E^{(0)} = R(0)$$

$$k_i = \left[R(i) - \sum_{j=1}^{i-1} \alpha_j^{i-1} R(i-j) \right] / E^{(i-1)} \qquad 1 \le i \le p$$

$$\alpha_i^{(i)} = k_i$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \qquad 1 \le j \le i-1$$

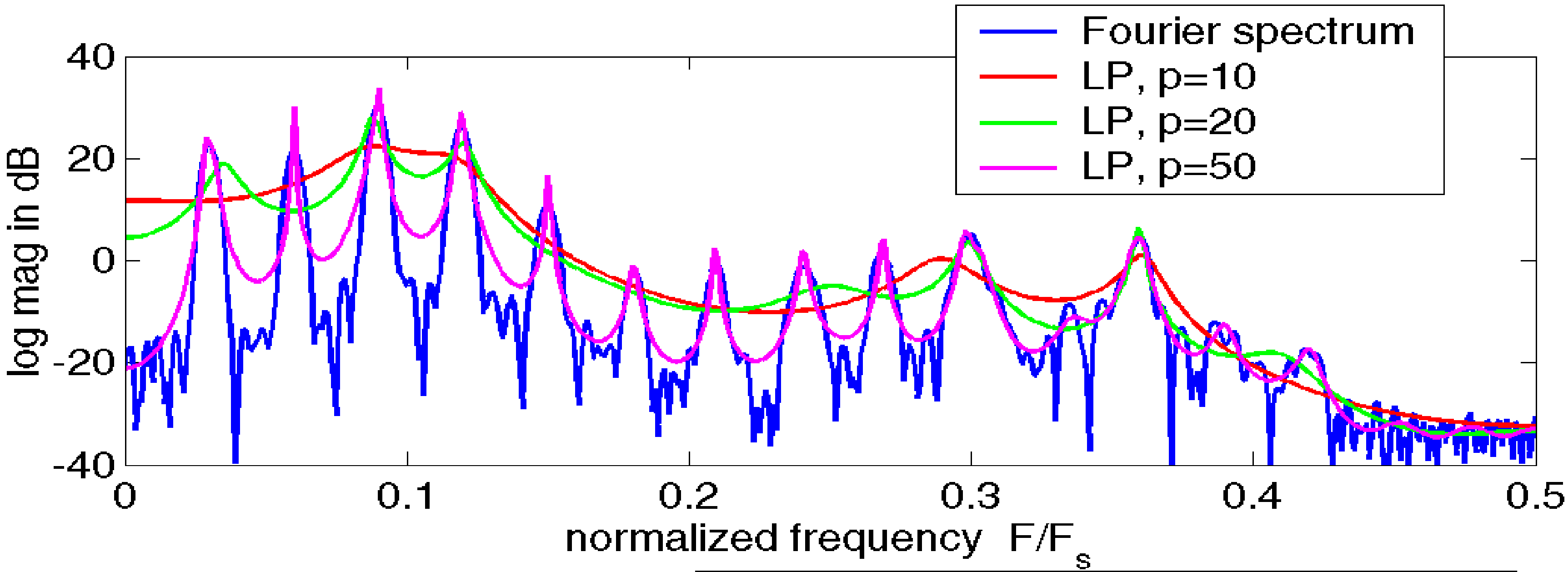
$$E^{(i)} = (1 - k_i^2) E^{(i-1)}$$

- solve equations recursively for $i = 1, 2, \ldots, p$ with the final solution

$$\alpha_j = \alpha_j^{(p)} \qquad 1 \le j \le p$$

- all predictors of orders $1, 2, 3, \ldots, p-1$ are found in Durbin's method, where $\alpha_j^{(i)}$ is the j^{th} predictor coefficient of an i^{th} order predictor

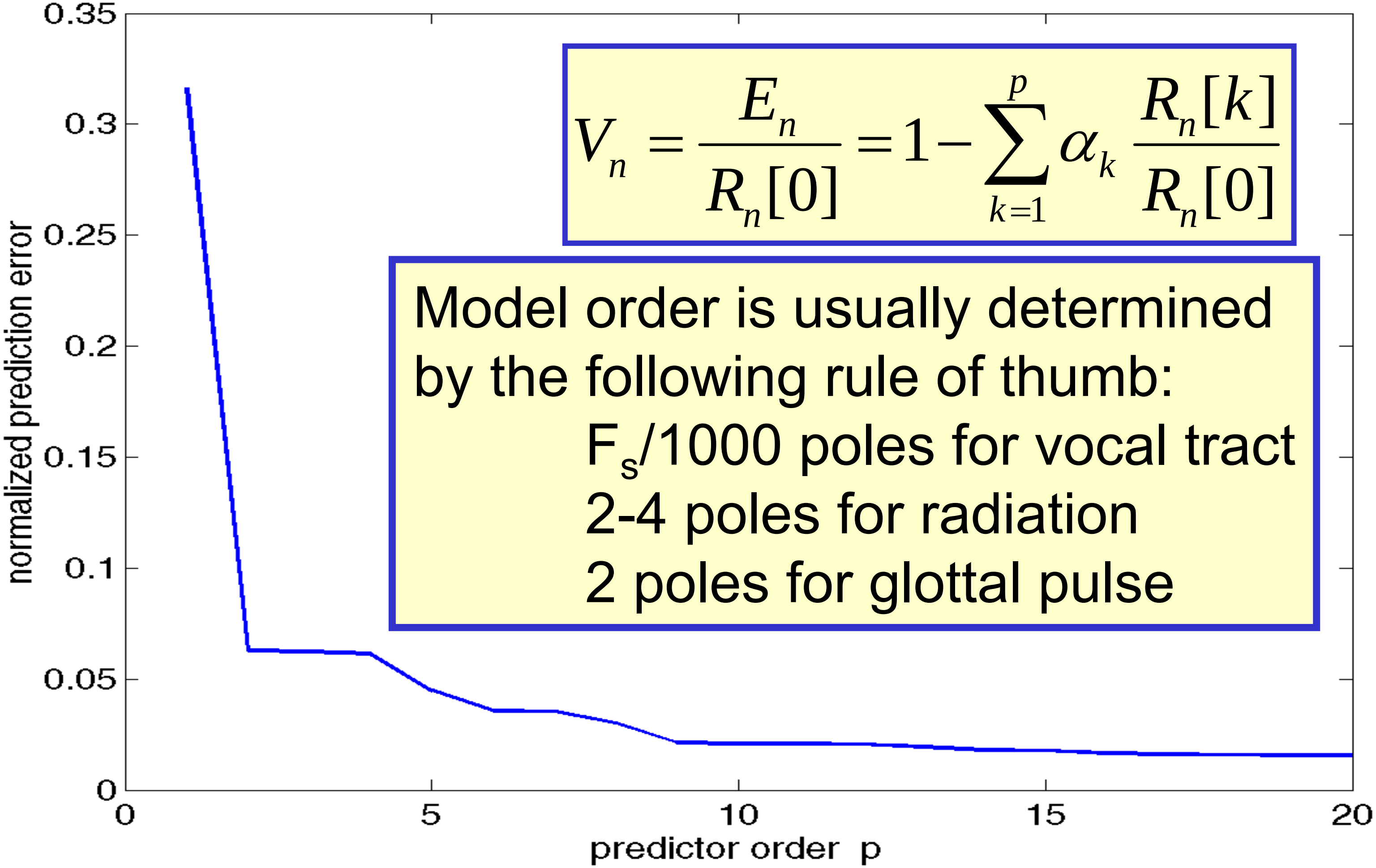
LPC Spectrum



$$H(e^{j\omega}) = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k e^{-j\omega k}}$$

```
x = s .* hamming(301);
X = fft( x , 1000 )
[ A , G , r ] = autolpc( x , 10 )
H = G ./ fft(A,1000);
```

Prediction Error as a Function of p



$$V_n = \frac{E_n}{R_n[0]} = 1 - \sum_{k=1}^{p} \alpha_k \frac{R_n[k]}{R_n[0]}$$

Model order is usually determined by the following rule of thumb:

- $F_s/1000$ poles for vocal tract
- 2-4 poles for radiation
- 2 poles for glottal pulse

Durbin's Algorithm (Levinson Recursion)

$E^0 = R[0]$

for $i = 1, 2, \ldots, p$

$$k_i = \left(R[i] - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R[i-j] \right) / E^{(i-1)}$$

$\alpha_i^{(i)} = k_i$

if $i > 1$, then for $j = 1, 2, \ldots, i-1$

$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$

end

$E^{(i)} = (1 - k_i^2) E^{(i-1)}$

end

$\alpha_j = \alpha_j^{(p)} \qquad j = 1, 2, \ldots, p$

The k_i are called PARCOR coefficients

$$\Rightarrow A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1})$$

The $E^{(i)}$ are the prediction error for an i^{th}-order predictor

Fallout from the Durbin Algorithm

- each iteration computes an i-th order predictor and the associated error, i.e.,

$$A^{(i)}(z) = 1 - \sum_{j=1}^{i} \alpha_j^{(i)} z^{-j}$$

$$E^{(i)} = R[0] - \sum_{j=1}^{i} \alpha_j^{(i)} R[j]$$

- it can be seen from the Durbin algorithm that

$$E^{(p)} = R[0] \cdot \prod_{i=1}^{p} (1 - k_i^2) > 0$$

$$-1 < k_i < 1$$

- the quantities k_i are called the PARCOR (partial correlation) coefficients. The above result implies

PARCORs to Prediction Coefficients

- assume that $k_i, i = 1, 2, \ldots, p$ are given. Then we can skip the computation of k_i in the Levinson recursion.

$$
\begin{aligned}
&\text{for } i = 1, 2, \ldots, p \\
&\quad \alpha_i^{(i)} = k_i \\
&\quad \text{if } i > 1, \text{ then for } j = 1, 2, \ldots, i-1 \\
&\qquad \alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \\
&\quad \text{end} \\
&\text{end} \\
&\alpha_j = \alpha_j^{(p)} \qquad j = 1, 2, \ldots, p
\end{aligned}
$$

Prediction Coefficients to PARCORs

- assume that $\alpha_j,\ j=1,2,\ldots,p$ are given. Then we can work backwards through the Levinson Recursion.

$$\alpha_j^{(p)} = \alpha_j \quad \text{for } j = 1,2,\ldots,p$$

$$k_p = \alpha_p^{(p)}$$

$$\text{for } i = p, p-1, \ldots, 2$$

$$\quad \text{for } j = 1,2,\ldots,i-1$$

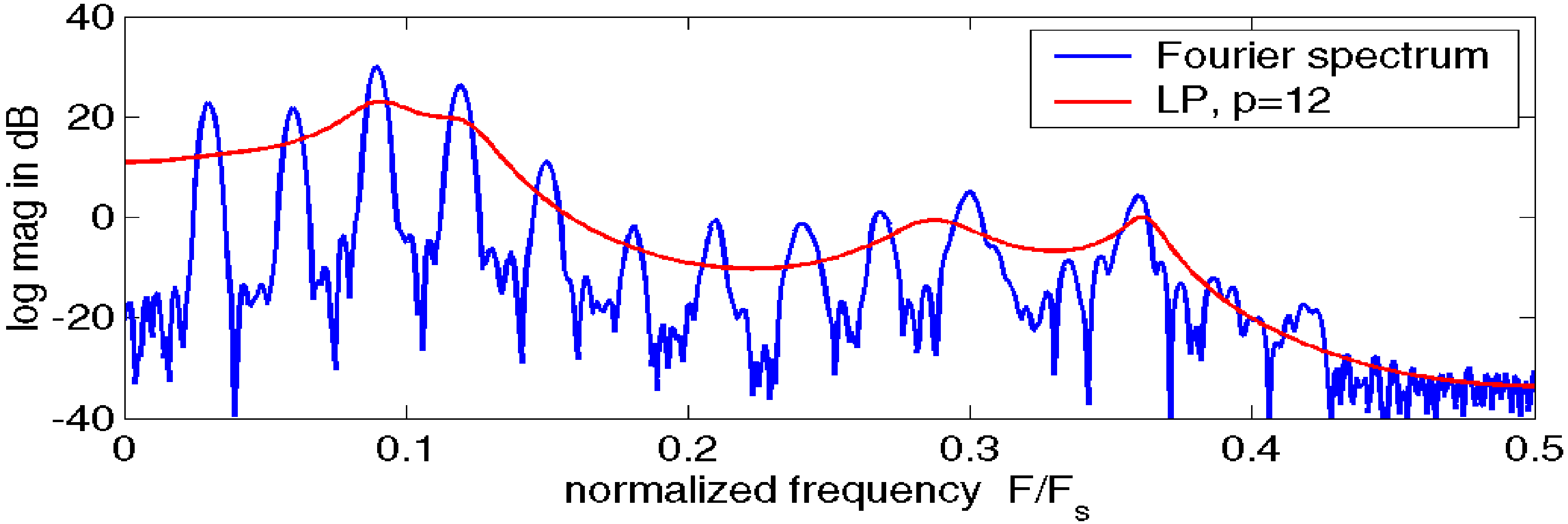
$$\qquad \alpha_j^{(i-1)} = \frac{\alpha_j^{(i)} + k_i \alpha_{i-j}^{(i)}}{1-k_i^2}$$

$$\quad \text{end}$$

$$\quad k_{i-1} = \alpha_{i-1}^{(i-1)}$$

$$\text{end}$$

12^{th}-Order Example



$$H(z) = \frac{G}{A(z)} = \frac{G}{1 - \sum_{i=1}^{p} \alpha_i z^{-i}} = \frac{G}{\prod_{i=1}^{p} (1 - z_i z^{-1})} == \frac{G z^p}{\prod_{i=1}^{p} (z - z_i)}$$

Minimum-Phase Property of *A*(*z*)

$A(z)$ has all its zeros inside the unit circle

Proof: Assume that z_i $(|z_i|^2 > 1)$ is a zero of $A(z)$

$$A(z) = (1 - z_i z^{-1}) A'(z)$$

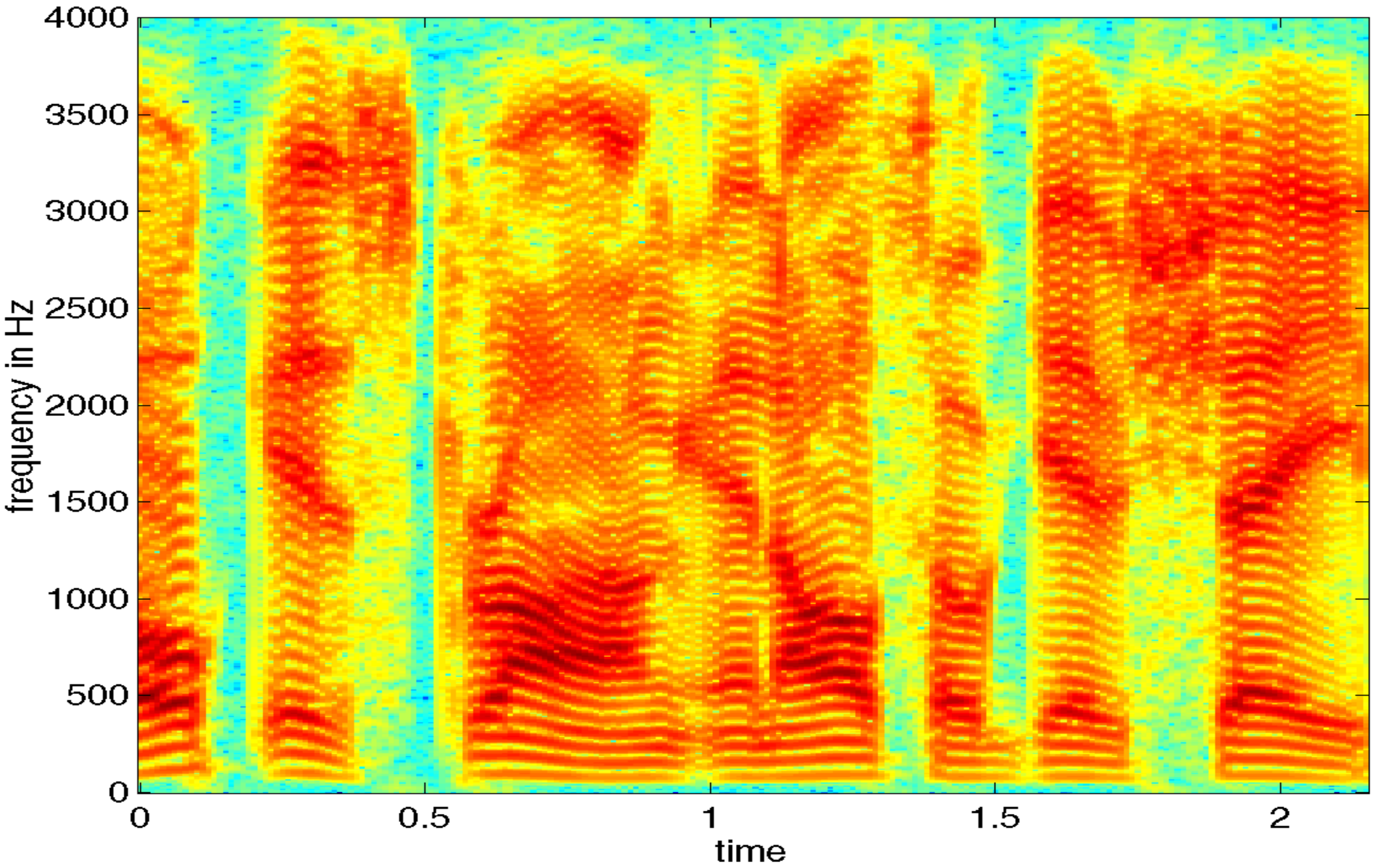
The minimum mean-squared error is

$$E_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|1 - z_i e^{-j\omega}\right|^2 \left|A'(e^{j\omega})\right|^2 \left|S_n(e^{j\omega})\right|^2 d\omega > 0$$

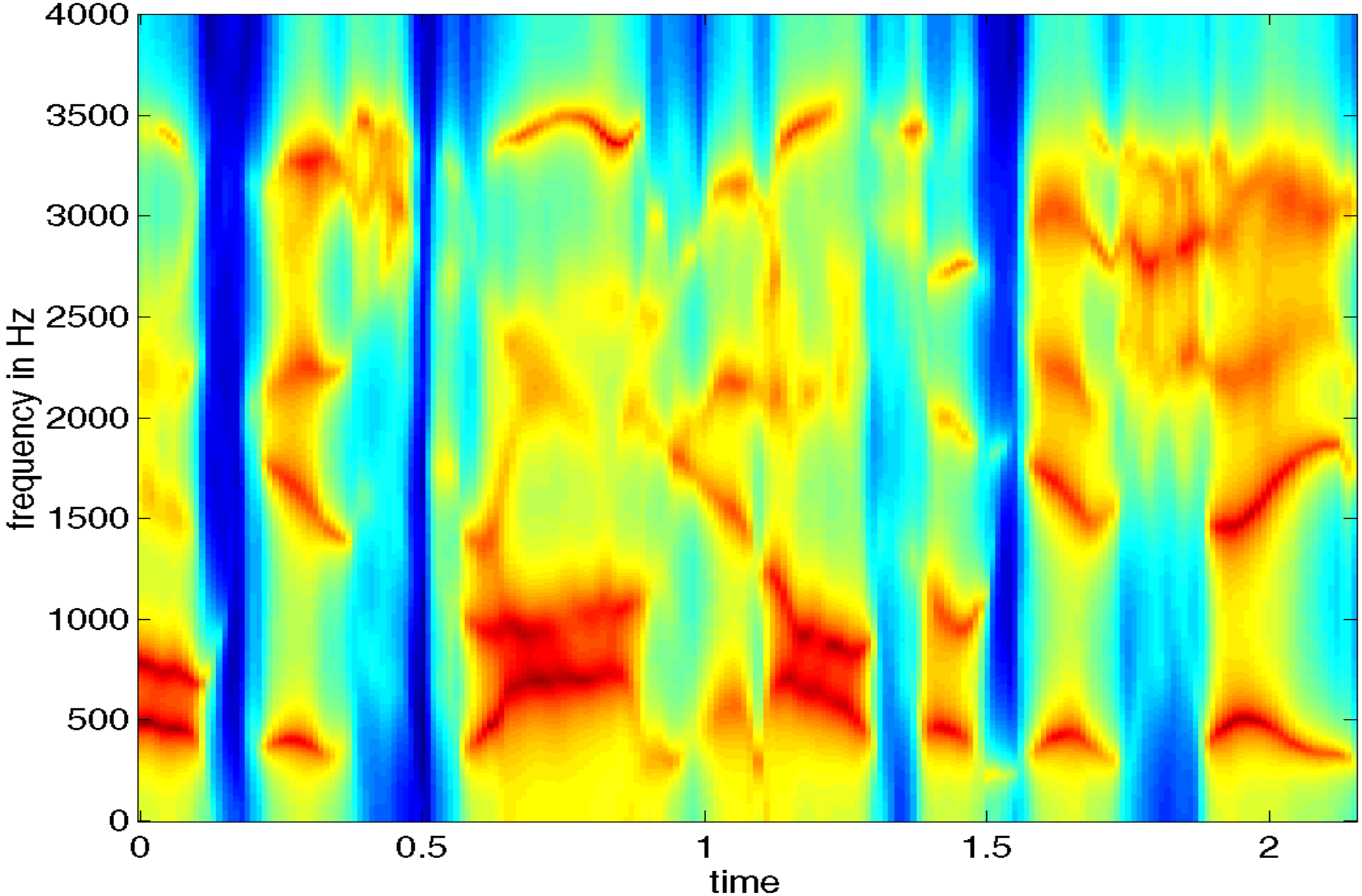
$$\left|1 - z_i e^{-j\omega}\right|^2 = \left|z_i\right|^2 \left|1 - (1/z_i^*) e^{-j\omega}\right|^2$$

Thus, $A(z)$ could not be the optimum filter because we could replace z_i by $(1/z_i^*)$ and decrease the error.

Narrowband Spectrogram



LP Spectrogram: An Example



LP Spectrogram

- the spectrogram is an image plot of the short-time Fourier transform

$$\left|X_r[k]\right| = \left|\sum_{m=0}^{N-1} x[rR+m]w[m]e^{-j(2\pi/N)km}\right|$$

- the LP spectrogram is a plot of

$$\left|H_r[k]\right| = \left|\frac{G_r}{A_r(e^{j(2\pi/N)k})}\right|$$

- how can you compute H from X? There is a direct path between them without involving the speech signal.

$$t_r = rRT$$

$$F_k = \frac{kF_s}{N}$$

PARCORs and Stability

- prove that $\left|k_i\right| \geq 1 \Rightarrow \left|z_j^{(i)}\right| \geq 1$ for some j

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) = \prod_{j=1}^{i}(1 - z_j^{(i)} z^{-1})$$

It is easily shown that $-k_i$ is the coefficient of z^{-i} in $A^{(i)}(z)$, i.e., $\alpha_i^{(i)} = k_i$. Therefore,

$$\left|k_i\right| = \prod_{j=1}^{p} z_j^{(i)}$$

If $\left|k_i\right| \geq 1$, then either all the roots must be **on** the unit circle or at least one of them must be **outside** the unit circle.

PARCORs and Stability

- if $|k_i| \geq 1,$ then either all the roots must be **on** the unit circle or at least one of them must be **outside** the unit circle. Since this is true for all $A^{(i)}(z)$, $i = 1, 2, \ldots, p,$ a necessary condition for the roots of $A^{(p)}(z)$ to be inside the unit circle is:

$$|k_i| < 1, \quad i = 1, 2, \ldots, p$$

- for the ith-order optimum linear predictor,

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} = \prod_{j=1}^{i} (1 - k_j^2) E^{(0)} > 0$$

so $|k_i| < 1$ and therefore $A^{(p)}(z)$ has all its roots inside the unit circle.

Summary

- This week's Classes
  - Linear method for regression (Chapter 3)
  - Autoregression and linear predictive speech analysis
- Next Class
  - Linear methods for classification
- Exercises: make sure you know the topics discussed and how to do all the exercises suggested in Lecture 2
- Reading Assignments
  - HTF, Chapter 4