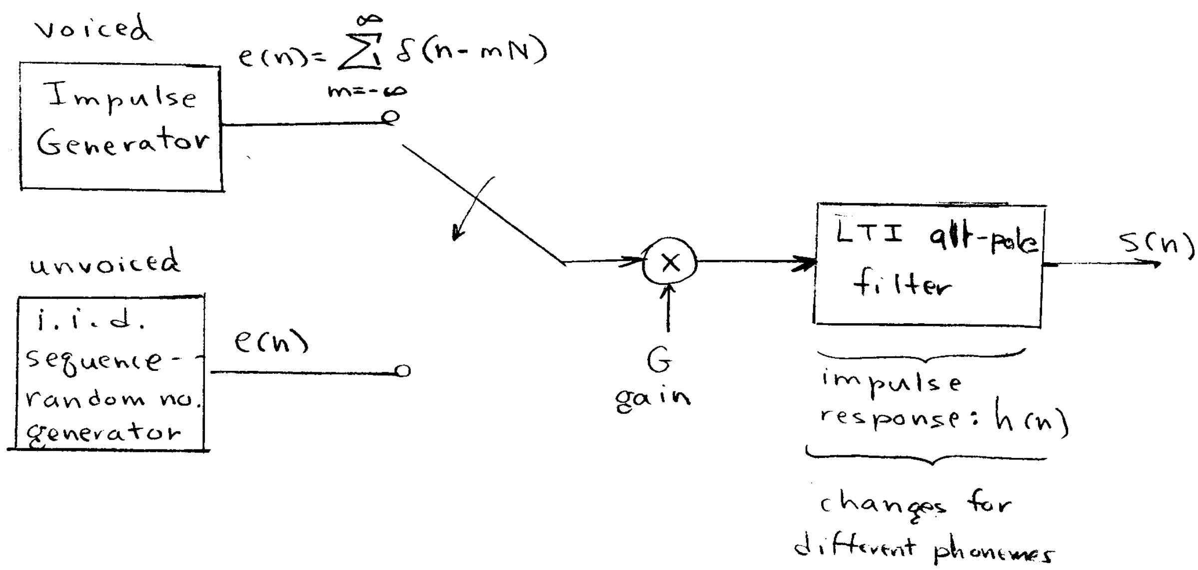
Lecture 22
Linear Predictive Coding (LPC)

• consider DT model of speech production:



• note: Keep in mind underlying sampling rate f_s:

$$NT_s = \frac{1}{f_0} \qquad \frac{N}{f_s} = \frac{1}{f_0} \Rightarrow f_s = N f_0$$

• G: accounts for variations in amplitude

• all-pole model: $$V(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{S(z)}{E(z)}$$

• model parameters: voiced/unvoiced classification
pitch period, N
gain G
coeffs. a_k

$z = e^{j\omega}$

• again: all-pole model is well-suited to resonant character of vocal tract in the case of non-nasal sounds

- angular locations of poles determines formants
- comment on hmwk. problem relating poles to " "

- nasal sounds are best modeled with some zeroes -- but estimation of coefficients of model then becomes difficult

- in practice, with enough poles get good estimate of both nasal & non-nasal sounds

  - in practice: ~ 13-14 poles for speech sampled at 10KHz

- <u>time-domain model</u>:

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

- <u>linear predictor</u>:

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

- <u>prediction error</u>:

$$f(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k \, s(n-k)$$

- pick α_k's, $k = 1, 2, \ldots, p$ to minimize $\sum_n f^2(n)$

- hopefully: α_k is an estimate of a_k

- let $A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$

- consider: $s(n) \longrightarrow \boxed{A(z)} \longrightarrow f(n)$

- if $\alpha_k = a_k$, $k = 1, \ldots, p$, then $V(z) = \dfrac{G}{A(z)}$

- in this case:

$$e(n) \longrightarrow \boxed{V(z)} \xrightarrow{s(n)} \boxed{A(z)} \xrightarrow{f(n)}$$

- if model was exact, and if coeff's were estimated correctly, then

$$f(n) = G\, e(n)$$

$\Rightarrow$ prediction error provides useful info. on pitch period

- in practice, we don't know G or a_k.

- have to estimate these quantities based on $s(n)$ and other a-priori information

- Estimation of LPC Coefficients

- we will estimate LPC coefficients for each "frame" sequentially
  - frame : windowed speech segment

- define short-time quantities:

$$s_n(m) = w(m)\, s(m+n)$$

  - window: $w(m) \neq 0$ only for $0 \leq m \leq N-1$

  - shift speech segment of interest to origin and window it

- $f_n(m) = s_n(m) - \hat{s}_n(m)$

  - note: $s_n(m) \neq 0$ only for $0 \leq m \leq N-1$

- $$\hat{s}_n(m) = \sum_{k=1}^{p} \alpha_k\, s_n(m-k)$$

  - note: $\hat{s}_n(m) \neq 0$ only for $0 \leq m \leq N+p-1$

  - thus: $f_n(m) \neq 0$ " "

- sum of square errors:

$$E_n = \sum_{m=0}^{N+p-1} f_n^2(m)$$

- Comments on limits;
- epochs in prediction error signal:

a) $0 \le m \le p-1 \Rightarrow$

$$\hat{S}_n(m) = \sum_{k=1}^{p} \alpha_k S_n(m-k) = \sum_{k=1}^{m} \alpha_k S_n(m-k)$$

- predicting $S_n(m)$ in terms of $m < p$ past values
- larger than normal error

b) $p \le m \le N-1 \Rightarrow$

- $S_n(m)$ is predicted in terms of p past values

c) $N \le m \le N+p-1 \Rightarrow$

- $S_n(m) = 0$ for $m > N-1 \Rightarrow$

  trying to predict $\emptyset$ from p past values
- larger than normal error

- If we restrict m to the normal range, least square error (LSE) solution yields <u>Covariance Method</u>
- Allowing m to vary over all three ranges, LSE approach yields <u>Autocorrelation Method</u>
- Why two methods?? ⇒ will discuss shortly
- at this point ⇒ just point out classic trade-off between performance vs. computational complexity

$$\Downarrow \qquad\qquad \Downarrow$$

Covariance Method vs. Autocorrelation Method

- We will first derive the Covariance Method
  - multi-dimensional unconstrained optimization problem:

$$\underset{\alpha_1, \alpha_2, \ldots, \alpha_p}{\text{Minimize}} \quad E_n = \sum_{m=p}^{N-1} f_n^2(m) = \sum_{m=p}^{N-1} \left\{ s_n(m) - \hat{s}_n(m) \right\}^2$$

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

- take partial derivative with respect to each α_i, $i = 1, 2, \ldots, p$, and set each equal to zero

· Covariance Method (cont.)

$$E_n = \sum_{m=p}^{N-1} f_n^2(m) = \sum_{m=p}^{N-1} \{ s_n(m) - \hat{s}_n(m) \}^2$$

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

$$\frac{\partial E_n}{\partial \alpha_\ell} = 2 \sum_{m=p}^{N-1} \left\{ s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right\} \left\{ 0 - s_n(m-\ell) \right\} = 0$$

$$= \sum_{m=p}^{N-1} s_n(m) s_n(m-\ell) = \sum_{m=p}^{N-1} \sum_{k=1}^{p} \alpha_k s_n(m-k) s_n(m-\ell)$$

$$= \sum_{k=1}^{p} \alpha_k \sum_{m=p}^{N-1} s_n(m-k) s_n(m-\ell)$$

$\ell = 1, 2, \ldots, p$

· define: $$\phi_n(\ell, k) \triangleq \sum_{m=p}^{N-1} s_n(m-k) s_n(m-\ell) = \phi(k, \ell)$$

$$\Rightarrow \phi_n(\ell, 0) = \sum_{k=1}^{p} \alpha_k \phi_n(\ell, k) \qquad \ell = 1, 2, \ldots, p$$

· p eqns. in p unknowns:

$$\begin{array}{c} \ell=1 \\ \ell=2 \\ \\ \\ \\ \end{array} \underbrace{\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & \phi_n(1,3) & \cdots & \phi_n(1,p) \\ \phi_n(2,1) & \phi_n(2,2) & \phi_n(2,3) & \cdots & \phi_n(2,p) \\ & & \vdots & & \\ \phi_n(p,1) & \phi_n(p,2) & \phi_n(p,3) & \cdots & \phi_n(p,p) \end{bmatrix}}_{\text{symmetric matrix since } \phi_n(k,\ell) = \phi_n(\ell,k)} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \vdots \\ \phi_n(2,p) \end{bmatrix}$$

- to solve for LPC coefficients via Covariance Method requires solution of linear system of p equations;
  - in matrix form:

$$\underset{p \times p}{\Phi} \ \underset{p \times 1}{\underline{\alpha}} = \underset{p \times 1}{\phi} \quad \Big\} \text{ where: } \Phi \text{ is symmetri}$$

SK·0

  - this must be solved for each windowed speech segment
  - * requires $O(p^3)$ operations

Lecture 23

- <u>Autocorrelation Method</u>:
  - in this case,

$$\underset{\alpha_1, \alpha_2, \ldots, \alpha_n}{\text{Minimize}} \quad E_n = \sum_{m=0}^{N+p-1} f_n^2(m)$$

$$= \sum_{m=0}^{m=0} \left\{ s_n(m) - \sum_{k=1} \alpha_k s_n(m-k) \right\}$$

  - recall previous observation $f_n(m) \neq 0$ only for $0 \leq m \leq N+p-1$.
  - Therefore, we may let limits in sum be infinite;

$$E_n = \sum_{m=-\infty}^{\infty} f_n^2(m)$$

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

- taking derivative wrt α_ℓ and equating to zero as before yields:

$$\sum_{m=-\infty}^{\infty} s_n(m)\, s_n(m-\ell) - \sum_{k=1}^{p} \alpha_k \sum_{m=-\infty}^{\infty} s_n(m-k)\, s_n(m-\ell) = \emptyset$$

$$\ell = 1, 2, \ldots, p$$

- note, due to limits:

$$\sum_{m=-\infty}^{\infty} s_n(m-k)\, s_n(m-\ell) \underset{\substack{\text{change of} \\ \text{variables:}}}{=} \sum_{m'=-\infty}^{\infty} s_n(m'+\ell-k)\, s_n(m')$$

$$m' = m - \ell \Rightarrow m' \Big]_{-\infty}^{\infty}$$

$$= \sum_{m=-\infty}^{\infty} s_n(m)\, s_n(m+\ell-k)$$ } only depends on difference, $\ell - k$ · not true for covariance method

- Define: $$R_n(\ell) \triangleq \sum_{m=-\infty}^{\infty} s_n(m)\, s_n(m+\ell)$$

- note: $$R_n(-\ell) = \sum_{m=-\infty}^{\infty} s_n(m)\, s_n(m-\ell)$$

· change of variables: $m' = m - \ell \Rightarrow m' \Big]_{-\infty}^{\infty}$

$$= \sum_{m'=-\infty}^{\infty} s_n(m'+\ell)\, s_n(m') = R_n(\ell)$$

- Hence, $$R_n(\ell) = \sum_{k=1}^{p} \alpha_k R_n(k-\ell)$$

$$\ell = 1, 2, \ldots, p$$

- defines p equations in p unknowns
- collectively:

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \cdots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \cdots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \cdots & R_n(p-3) \\ \vdots & \vdots & \vdots & & \vdots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \cdots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \vdots \\ R_n(p) \end{bmatrix}$$

- in matrix form: $\underset{p \times p}{\underline{R}_n} \, \underset{p \times 1}{\underline{\alpha}} = \underset{p \times 1}{\underline{r}_n}$.

- observe properties of $\underline{R}_n$:

  1. symmetric about main diagonal
     - equal to its own transpose
  2. constant along any diagonal
     - referred to as Toeplitz

- Thus, the system of equations obtained via the Autocorrelation Method is Toeplitz as well as symmetric

- structure may be exploited to efficiently solve for LPC coefficients

- Note: it can be shown (homework):

$$E_n = \sum_{m=0}^{N+p-1} \left\{ s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right\}^2$$

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

- Practical Considerations:

2. A window with taper is typically employed (once again -- Hamming) to counteract boundary effects with Autocorrelation method -- don't need window for Cov. Method

1. Accuracy of method -- covariance method wins!
   Computation: Covariance method: $O(p^3)$
   Autocorrelation " ": $O(p^2)$
   - need to solve matrix eqns. for each "frame"

3. Stability: $V(z) = \dfrac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \dfrac{1}{A(z)}$

   - Autocorrelation method: guaranteed roots of $A(z)$ are within unit circle -- stable
   - Covariance method: possible for roots of $A(z)$ to be outside unit circle

- Autocorrelation method used in practice

4. How many poles do we need?

$$1 \frac{\text{formant}}{\text{KHz}} \cdot \underbrace{\frac{2 \text{ poles}}{\text{formants}}}_{\text{complex conjugate pair}} \cdot \frac{F_s}{2} = \#\text{ of poles}$$

   - for 10 KHz => 10 poles

- radiation from mouth - 1 pole
- glottal pulse shape - 1 pole
- to compensate for nasals, add some extra poles

- for speech sampled at $F_s = 10$ KHz : 13-15 poles

5. How long should window be?
   - want vocal tract configuration fixed
   - need to include several pitch periods to eliminate effect of window taper
   - typically : 10-40 ms

6. LPC Performance -- see Figures in handout

7. Transmission bit rate:
   - Typical LPC Vocoder Parameters:
     - p=14 coefficients } works well = 84 bits/frame
     - 6 bits / coefficient }

84 bits / frame
1 bit => voiced/ unvoiced
6 bits => pitch (N)
5 bits => Gain (G)
96 bits / frame

for 50 frames/sec : 4800 bps

for 100 frames/sec : 9600 bps

- recall : point of reference : code digitized speech directly
  - telephone quality : $f_s \approx 6$ KHz, # of bits = 7 (SNR ≈ 36 d
    => 42,000 bps
  - high quality : $F_s = 20$ KHz, # of bits = 11 (SNR ≈ 60 dB
    => 220,000 bps

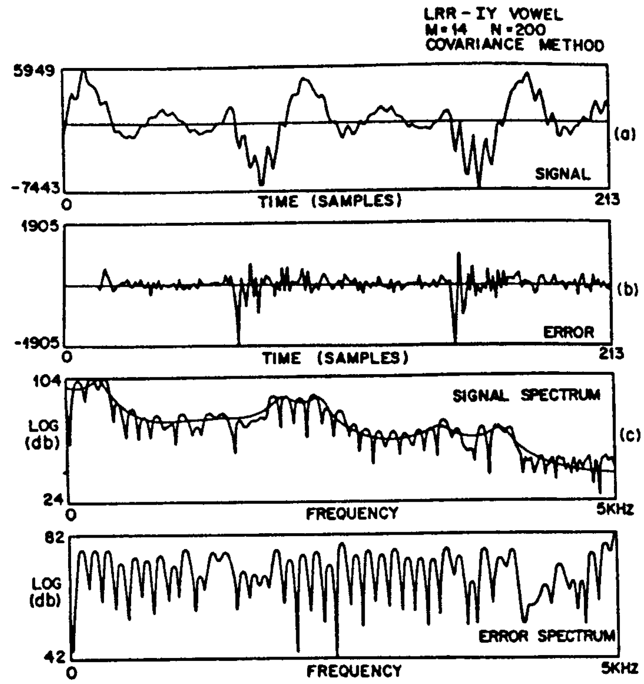


Fig. 8.6 Typical signals and spectra for LPC covariance method for a male speaker. (After Rabiner et al. [16].)

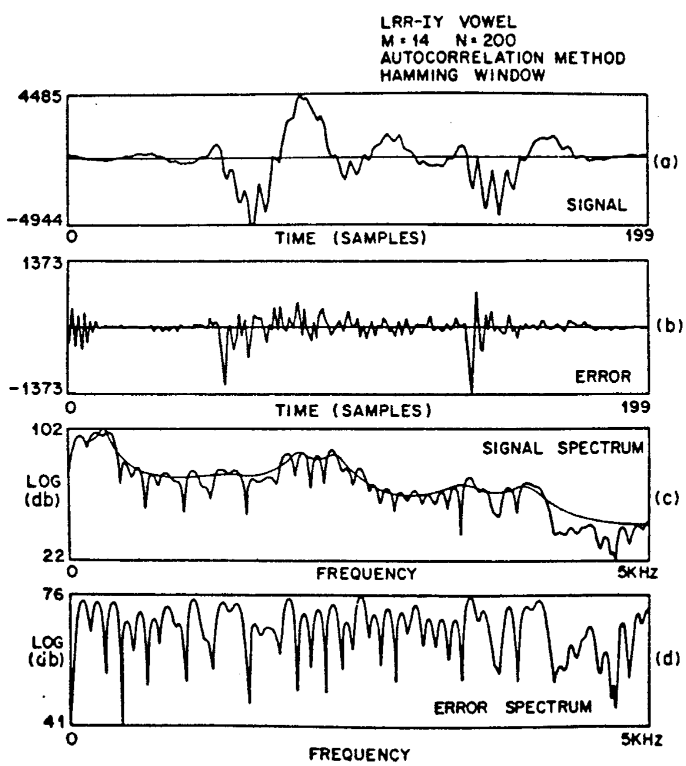
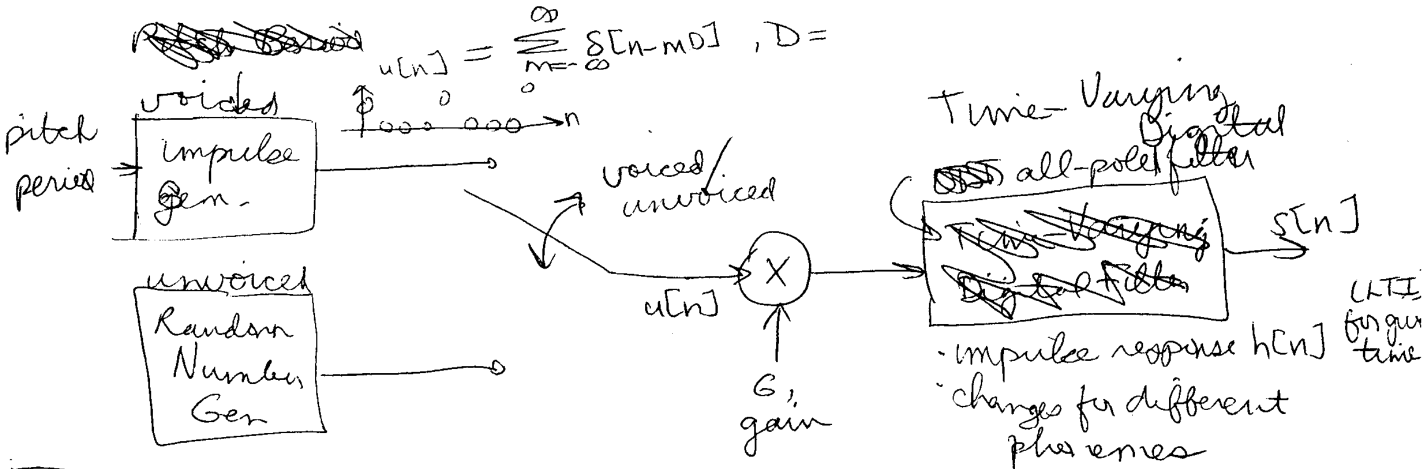


Fig. 8.7 Typical signals and spectra for LPC autocorrelation method for a male speaker. (After Rabiner et al. [16].)

LPC ~~of~~ Speech (Linear Predictive Coding of Speech) 5

– estimate pitch, formants, spectra, vocal tract function

- Consider Model:



Model

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

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

Model Parameters

voiced / unvoiced classification

pitch period, D

gain G

coefficients a_k

Extra

The all-~~pole~~ pole model is well-suited to resonant character of vocal tract in the case of non-nasal sounds.

estimate nasal sounds w/ some zeroes – but this is difficult

In practice, with enough poles, get good estimate of both nasal and non-nasal sounds.

Time-domain Model ← all parameters assumed to vary slowly w/ time!

- $s[n] = \sum_{k=1}^{p} a_k s[n-k] + G\,u[n]$

Linear Predictor – predict coefficients of the model

$$\hat{s}[n] = \sum_{k=1}^{p} \alpha_k s[n-k]$$

^ predicted speech

Prediction Error:

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

Let $A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$.

- Then have

$$s[n] \rightarrow \boxed{A(z)} \rightarrow e[n]$$

Note ~~that~~ if $\alpha_k = a_k$, $k = 1, \ldots, p$, then

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

So we have

$$u[n] \rightarrow \boxed{H(z)} \xrightarrow{s[n]} \boxed{A(z)} \rightarrow e[n]$$

If (model were exact, and if coeff. were estimated exactly) $\alpha_k = a_k$,

- then $e[n] = G u[n]$. ⇒ Prediction error (provides useful information about the pitch period.) contains pitch information.

In practice, we don't know G or a_k.

We have to <u>estimate</u> these quantities based on $s[n]$ and other <u>apriori</u> information.

Choose prediction coefficients to minimize mean-squared error.

MMSE - Minimum Mean Square Estimate

(1) For assumed system model, $a_k = \alpha_k$ is a solution
(2) Based on $e[n] = G\,u[n]$, error will be small
(3) Computationally nice - set of linear equations

extra

Minimize $\sum_n e^2[n]$.

~~Est~~ Estimate coefficients for a windowed speech segment.

~~Define~~ Short-time Quantities definitions:

$s_n[m] = w[m]\,s[m+n]$

$w[m] \neq 0$ for $0 \le m \le N-1$

(Shift speech segment of interest to origin and window.)

error: $e_n[m] = s_n[m] - \tilde{s}_n[m]$

$$\tilde{s}_n[m] = \sum_{k=1}^{p} \alpha_k s_n[m-k]$$

- linear predictor
- not zero until $m = N+p$.

Sum of Squared error:

$$E_n = \sum_{m=0}^{N+p-1} e_n^2[m]$$

Error will be larger near endpoints of $e_n[m]$ because

$0 \le m < p$: predict $s_n[m]$ in terms of $m < p$ past values

- $N \le m \le N+p$: $s_n[m] = 0$, trying to predict $s_n[m]$ from p past values

Make window ~~taper~~ attenuate points at ends to keep error down at edges of $\tilde{s}_n[m]$

Covariance Method

$$E_n = \sum_{m=p}^{N-1} e_n^2[m]$$ ← ~~Not as good~~ Better performance

- We won't do this.

Autocorrelation Method

$$E_n = \sum_{m=-\infty}^{\infty} e_n^2[m]$$ ← ~~More~~ Less computationally complex.

$$= \sum_m \left[s_n[m] - \tilde{s}_n[m] \right]^2$$

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

Minimize E_n.

- Differentiate with respect to α_ℓ. (ℓ-fixed)

$$\frac{\partial E_n}{\partial \alpha_\ell} = \sum_m 2\left[s_n[m] - \sum_{k=1}^{p} \alpha_k s_n[m-k] \right] \frac{\partial}{\partial \alpha_\ell} \left[s_n[m] - \sum_{k=1}^{p} \alpha_k s_n[m-k] \right]$$

$$= \sum_m 2\left[s_n[m] - \sum_{k=1}^{p} \alpha_k s_n[m-k] \right] \left[- s_n[m-\ell] \right]$$

- Set $\frac{\partial E_n}{\partial \alpha_\ell} = 0$

$$\sum_m s_n[m] s_n[m-\ell] - \sum_{k=1}^{p} \alpha_k \sum_m s_n[m-k] s_n[m-\ell] = 0 \quad ; \quad \ell = 1, 2, \dots, p$$

Define $R_n[k] = \sum_m s_n[m] s_n[m+k]$ This is autocorrelation.

We can [illegible]

~~note: We can say this~~

So,

$$R_n[\ell] - \sum_{k=1}^{p} \alpha_k R_n[k-\ell] = 0 \quad , \ell = 1, 2, \dots, p$$

aside: $\sum_m s_n[m-k] s_n[m-\ell]$

$n = m-k \Rightarrow m = n+k$

$\sum_n s_n[n] s_n[n+k-\ell]$

$= R_n[k-\ell]$

Note: We can say this because of the infinite limit.

Collectively, we can write the set of p linear equations as a matrix-vector product.

$$\underbrace{\begin{bmatrix} R_n(0) & R_n(1) & \cdots & R_n(p-1) \\ R_n(1) & R_n(0) & \cdots & R_n(p-2) \\ \vdots & & & \\ R_n(p-1) & R_n(p-2) & \cdots & R_n(0) \end{bmatrix}}_{\underline{R}} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}}_{\underline{\alpha}} = \underbrace{\begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix}}_{\underline{r}}$$

$\underline{R}$ properties:

① Symmetric

$$[\underline{R}]_{ij} = [\underline{R}]_{ji}$$

② constant along any diagonal

$$[\underline{R}]_{ij} = [\underline{R}]_{i+k, j+k}$$

$\Rightarrow \underline{R}$ is a Toeplitz matrix.

LPC

To solve for unknown coefficients, invert $\underline{R}$.

- direct inversion (Gaussian elimination) $\Rightarrow p^3$ operations
- fast method which exploits structure of $\underline{R} \Rightarrow p^2$ operations
  - Levinson – Durbin recursion

Practical Considerations

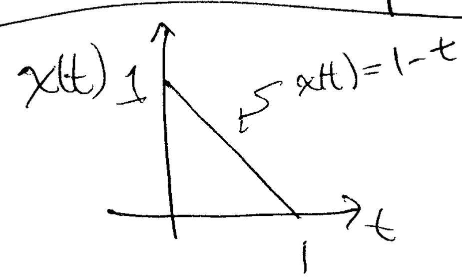
① Use smoothly tapered window to minimize boundary effects

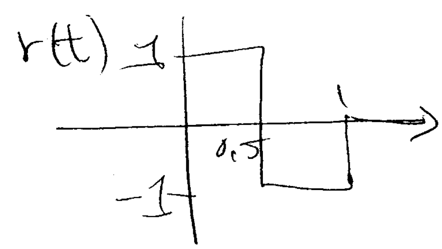
② Stability: Poles guaranteed to be inside unit circle using autocorrelation method
- not for covariance method.

③ Need 1 pole per kHz of sampling frequency, plus 3-4 poles due to ~~vocal tract glott~~ source excitation and radiation compensation

④ Include several pitch periods in the window, e.g. $N \approx 400$ samples for $F_s = 10\text{kHz}$, $P = \frac{1}{100\text{Hz}}$

MMSE Example:





$$\hat{x}(t) = a_0 + a_1 r(t)$$

$$E_n = \int_0^1 [\hat{x}(t) - x(t)]^2 dt = \int_0^1 [a_0 + a_1 r(t) - x(t)]^2 dt$$

$$\frac{\partial E_n}{\partial a_0} = \int_0^1 2[a_0 + a_1 r(t) - x(t)]dt \equiv 0$$

$$\int_0^1 a_0\,dt + \int_0^1 a_1 r(t) - \int_0^1 x(t)\,dt = 0$$

$$a_0 + 0 - \frac{1}{2} = 0 \Rightarrow \boxed{a_0 = \frac{1}{2}} \qquad (1)$$

$$\frac{\partial E_n}{\partial a_1} = \int_0^1 2[a_0 + a_1 r(t) - x(t)]\, r(t)\,dt = 0$$

$$\int_0^1 a_0 r(t)\,dt + \int_0^1 a_1 r^2(t)\,dt - \int_0^1 x(t) r(t)\,dt = 0$$

Extra

$$0 + a_1 - \left[\int_0^{\frac{1}{2}} (1-t)\,dt + \int_{\frac{1}{2}}^1 (t-1)\,dt\right] = 0$$

$$a_1 - \left[\left(t - \frac{t^2}{2}\right)\Big|_0^{\frac{1}{2}} + \left(\frac{t^2}{2} - t\right)\Big|_{\frac{1}{2}}^{1}\right] = 0$$

$$a_1 - \left[\frac{1}{2} - \frac{1}{8} + \frac{1}{2} - 1 - \frac{1}{8} + \frac{1}{2}\right] = 0$$

$$a_1 - \frac{1}{4} = 0 \Rightarrow \boxed{a_1 = \frac{1}{4}} \qquad (2)$$

OR:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix}$$

$$\Rightarrow \quad a_0 = \frac{1}{2}$$

$$a_1 = \frac{1}{4}$$

$$\hat{x}(t) = \frac{1}{2} + \frac{1}{4} r(t)$$