ECE 438 Lecture Monday 10 April 2023

Announcements:

* There will be a quiz today!
  - Available on Gradescope at 6:00p EDT
  - You will 30 minutes to work your solution.
  - You must upload your solution by 11:59p EDT

* Office Hours today:
  - 2:30p EDT
  - 4:00p EDT

LPC of speech

Divide speech signal into frames

Let $s_n[m] = w[m]s[m+n]$
(n ← frame location)

$w[m]$ - window $\neq 0$ only for $0 \leq m \leq N-1$

predictor:

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

note that $s_n[m] \neq 0$ only for $0 \leq m \leq N+p-1$

Derived optimal coefficients that minimize total squared error

$$E_n = \sum_{m=-\infty}^{\infty} f_n^2[m]$$

define $R_n[l] = \sum_{m=-\infty}^{\infty} s_n[m] s_n[m+l]$

solution:

$$R_n[l] = \sum_{k=1}^{p} \alpha_k R_n[l-k] \quad l = 1, \ldots, p$$

$R_n[l-k] = R_n[k-l]$

$$\underline{R}\vec{\alpha} = \vec{r}$$

$$\vec{r} = \begin{bmatrix} R_n[1] \\ \vdots \\ R_n[p] \end{bmatrix} \quad \vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} \quad \underline{R} = \begin{bmatrix} R_n[0] & R_n[1] & \ldots & R_n[p-1] \\ R_n[1] & R_n[0] & R_n[1] \ldots & R_n[p-2] \\ & \ddots & & \\ R_n[p-1] & R_n[p-2] & \ldots & R_n[0] \end{bmatrix}$$

Toeplitz matrix

efficient solution: Levinson-Durbin recursion $O(p^2)$

autocorrelation method

Practical considerations

1.
2.
3.
4. How many poles do we need?

$$\left(\frac{\text{formant}}{\text{kHz}}\right)\left(\frac{2 \text{ poles}}{\text{formant}}\right)\left(\frac{F_s}{2}\right)$$

for 10 kHz speech, need 10 poles

(This is just modeling the vocal tract

additional poles:

radiation from mouth — 1 pole

glottal pulse — 1 pole

nasal sounds — several extra poles

⇒ 13–15 poles for 10kHz speech

(5) Window length

- short enough to get just one phoneme
- long enough to capture several pitch periods
- long enough to minimize effect of window taper

example: 10–40 ms. for 10k Hz speech ⇒

100 to 400 samples

typical pitch = 10 msec ⇒ 100 Hz pitch frequency

Practical considerations (cont.)

Need 13–15 poles for 10kHz speech

assume $p = 14$

6 bits/coeff. ⇒ 32 levels of magnitude

⇒ 84 bits/frame

1 bit unvoiced/voiced

6 bits pitch period

5 bits gain

= 96 bits/frame

at 50 frames/sec $\Rightarrow$ 4800 bits/sec

100 frames/sec $\Rightarrow$ 9600 bits/sec

Compare with direct waveform encoding

"toll quality speech"

$F_s = 6$ kHz at 7 bits/sample $\Rightarrow$ 42,000 bits/sec

SNR = 36 dB according to Prof. Zoltowski

"high fidelity speech"

$F_s = 20$ kHz at 11 bits/sample

SNR = 60 dB

$\Rightarrow$ 220,000 bits/sec

Ignore pp. 14-18 by former student James Owens

Final topic for Module 4.3:

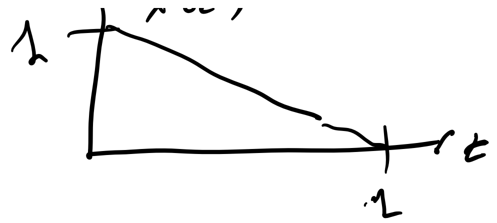
Least squares approximation of functions

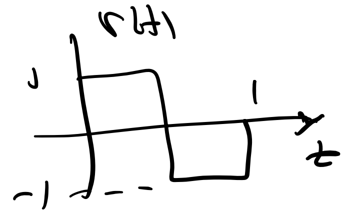
ideas are due to Robert N. McDonough

Will re-upload Module 4.3 with material on Rabiner and McDonough later

Simple example:

$x(t)$





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

2 basis function

Find a_0 & a_1 to minimize

$$E = \int_0^1 (\hat{x}(t) - x(t))^2 \, dt$$

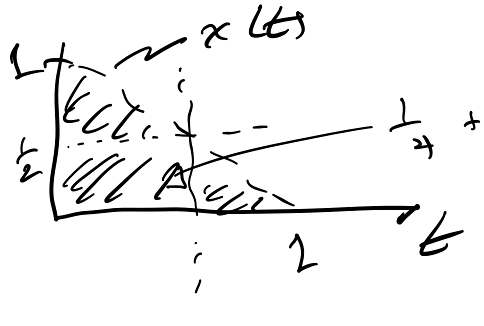
$$\frac{\partial E}{\partial a_0} = \int_0^1 \frac{\partial}{\partial a_0} (\hat{x}(t) - x(t))^2 \, dt$$

$$= \int_0^1 2(\hat{x}(t) - x(t)) \frac{\partial}{\partial a_0}(\cdot) \, dt$$

$$\frac{\partial}{\partial a_0}(\cdot) = \frac{\partial}{\partial a_0}[a_0 + a_1 r(t)] = (1)$$

$$\int_0^1 (\overset{(2)}{\hat{x}} - x(t))(1) \, dt = 0$$

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



$\frac{1}{4} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{8} = a_0 = \frac{1}{2}$

$$\frac{\partial}{\partial a_1}(E) = \int_0^1 2(\hat{x}(t) - x(t)) \frac{\partial}{\partial a_1}(\cdot)\, dt$$

$$\frac{\partial}{\partial a_1}(\cdot) = \frac{\partial}{\partial a_1}\{a_0 + a_1 r(t) - x(t)] = r(t)$$

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

$$= 0 + a_1 - \int x(t) r(t)\, dt \qquad \underline{a_1 = \frac{1}{4}}$$