Linear Prediction

- An idea of time-varying signal
  - Signal parameters change over time
- Lack of mathematical tools for dealing with 'time-varying' signals.
- Practical approach : time-invariant for a short period.
- frame-by-frame analysis of speech signal.

$x(k)$: is a speech signal for a short frame (for example, lets say 20 ms)

$\hat{x}(k)$: is the predicted signal

Linear prediction : $\hat{x}(k) = \sum_{i=1}^{m} a_i \, x(k-i)$.

$d(k)$: is the prediction error

$$d(k) = x(k) - \hat{x}(k)$$

using square error measure, we want to minimize

$$\mathcal{E}\{d^2(k)\} = \mathcal{E}\{(x(k) - \hat{x}(k))^2\}$$

objective : Learing (estimating) of $\{a_i\}_{i=1}^{m}$ LP coefficients

Assume $x(k)$ is a zero mean signal, $\mathcal{E}\{x(k)\}=0$.

$$\therefore \quad \mathcal{E}\{\hat{x}(k)\}=0, \quad \mathcal{E}\{d(k)\}=0$$

$$\sigma_d^2 = \mathcal{E}\{d^2(k)\} \quad : \text{ variance of } d(k).$$

$$\sigma_x^2 = \mathcal{E}\{x^2(k)\} \quad : \text{ variance of } x(k).$$

Basically, using LP, we remove correlations between speech samples. This is nothing but redundancy removal and acting upon the information content of the speech signal. So, we only need to code the relevant information and send to the user (receiver). We must need $\sigma_d^2 < \sigma_x^2$ for useful LP.

LP formulation (we want to minimize $\mathcal{E}\{d^2(k)\}$ and find optimal $\{a_i\}$).

$$\frac{\partial \mathcal{E}\{d^2(k)\}}{\partial a_\lambda} = 0, \quad \lambda = 1, 2, \cdots, m.$$

or, $\mathcal{E}\left\{2\, d(k) \frac{\partial d(k)}{\partial a_\lambda}\right\} = 0$; $\quad d(k) = x(k) - \hat{x}(k) = x(k) - \sum_{i=1}^{m} a_i\, x(k-i)$.

or, $\mathcal{E}\{-2\, d(k)\, x(k-\lambda)\} = 0$

or, $\mathcal{E}\{d(k)\, x(k-\lambda)\} = 0$

or, $\mathcal{E}\left\{\left[x(k) - \sum_{i=1}^{m} a_i\, x(k-i)\right] x(k-\lambda)\right\} = 0$

or, $\mathcal{E}\{x(k)\, x(k-\lambda)\} = \sum_{i=1}^{m} a_i\, \mathcal{E}\{x(k-i)\, x(k-\lambda)\}$

Assuming $x(k)$ is second order stationary,

the auto-correlation $\phi_{xx}(\lambda - i) = \mathcal{E}\{x(k-i)\, x(k-\lambda)\}$

$$\boxed{\therefore \quad \sum_{i=1}^{m} a_i\, \phi_{xx}(\lambda - i) = \phi_{xx}(\lambda), \quad \lambda = 1, 2, \cdots, m}$$

$$\underbrace{\begin{bmatrix} \phi_{xx}(0) & \phi_{xx}(-1) & \cdots\cdots & \phi_{xx}(1-m) \\ \phi_{xx}(1) & \phi_{xx}(0) & \cdots\cdots & \phi_{xx}(2-m) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{xx}(n-1) & \phi_{xx}(n-2) & \cdots\cdots & \phi_{xx}(0) \end{bmatrix}}_{R_{xx}} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}}_{\underline{a}} = \underbrace{\begin{bmatrix} \phi_{xx}(1) \\ \phi_{xx}(2) \\ \vdots \\ \phi_{xx}(m) \end{bmatrix}}_{\underline{\phi}_{xx}}$$

Auto-correlation matrix | LPC | Auto-correlation vector

$$\therefore \boxed{R_{xx}\,\underline{a} = \underline{\phi}_{xx}} \longrightarrow \text{Normal Equations}$$

or, $\underline{a}_{opt} = R_{xx}^{-1}\,\underline{\phi}_{xx}$.

Let we define

$$\underline{x}(k-1) = \begin{bmatrix} x(k-1) \\ x(k-2) \\ \vdots \\ x(k-m) \end{bmatrix}$$

NOW, $$\sigma_d^2 = \mathcal{E}\left\{\left(x(k) - \hat{x}(k)\right)^2\right\}$$

$$= \mathcal{E}\left\{ x^2(k) - 2x(k)\,\hat{x}(k) + \hat{x}^2(k) \right\}$$

$$= \mathcal{E}\left\{ x^2(k) - 2x(k).\,\underline{a}^T\underline{x}(k-1) + \left(\underline{a}^T\underline{x}(k-1)\right)^2 \right\}$$

$$\left[\because \hat{x}(k) = \sum_{i=1}^{m} a_i\,x(k-i) = \underline{a}^T\underline{x}(k-1)\right]$$

$$= \mathcal{E}\{x^2(k)\} - 2\,\underline{a}^T\,\mathcal{E}\{x(k)\,\underline{x}(k-1)\} + \underline{a}^T\,\mathcal{E}\{\underline{x}(k-1)\,\underline{x}^T(k-1)\}\,\underline{a}$$

$$= \sigma_x^2 - 2\,\underline{a}^T\underline{\phi}_{xx} + \underline{a}^T R_{xx}\underline{a}$$

$$= \sigma_x^2 - 2\underline{a}^T\underline{\phi}_{xx} + \underline{a}^T\underline{\phi}_{xx} \quad \left[\text{as } R_{xx}\underline{a} = \underline{\phi}_{xx}\right]$$

$$= \sigma_x^2 - \underline{a}^T\underline{\phi}_{xx}$$

$$= \sigma_x^2 - \sum_{\lambda=1}^{m} a_\lambda\,\phi_{xx}(\lambda) \longrightarrow \text{what } \sigma_d^2 \text{ we can have?}$$

or

$$= \sigma_x^2 - 2\underline{a}^T R_x\underline{a} + \underline{a}^T R_{xx}\underline{a}$$

$$= \sigma_x^2 - \underline{a}^T R_{xx}\underline{a} = \sigma_x^2 - \left(R_{xx}^{-1}\underline{\phi}_{xx}\right)^T R_{xx}\left(R_{xx}^{-1}\underline{\phi}_{xx}\right)$$

$$\boxed{\sigma_d^2 = \sigma_x^2 - \underline{\phi}_{xx}^T\left(R_{xx}^{-1}\right)^T\underline{\phi}_{xx}} \quad \left[\because \underline{a}_{opt} = R_{xx}^{-1}\underline{\phi}_{xx}\right]$$

- An idea of quantification, just dependent on signal properties.
- If we use optimal LPC with order 'm', then we can achieve this.

R_{xx} is an auto-correlation matrix. It will be positive semi-definite.

$$\therefore \phi_{xx}^T R_{xx}^{-1} \phi_{xx} \geq 0$$

Now, $\sigma_d^2 = \sigma_x^2 - \phi_{xx}^T R_{xx}^{-1} \phi_{xx}$

We know, $\sigma_d^2 \geq 0$, $\sigma_x^2 \geq 0$ and $\phi_{xx}^T R_{xx}^{-1} \phi_{xx} \geq 0$.

$$\therefore \boxed{\sigma_d^2 \leq \sigma_x^2}$$

Prediction gain: $G_p = \dfrac{\sigma_x^2}{\sigma_d^2}$

The prediction gain is a measure for the bit rate reduction that can be achieved by LP coding techniques

Spectral Flatness Measure

- $$\gamma_x^2 = \frac{\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left(\Phi_{xx}(e^{j\Omega})\right) d\Omega\right)}{\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{xx}(e^{j\Omega})\, d\Omega}.$$

Here $\Phi_{xx}(e^{j\Omega})$ is the power spectral density (PSD) of a signal $x(k)$.

If $\Phi_{xx}(e^{j\Omega}) = \text{const.} = \sigma_x^2$, we obtain $\gamma_x^2 = 1$

- γ_x^2 is nothing but $\frac{GM}{AM}$ of $\Phi_{xx}(e^{j\Omega})$.
- PSD $\xrightleftharpoons{FT}$ Auto-correlation $\{\phi_{xx}(k)\}$

NOW, using periodogram method, PSD is computed by the squared magnitude of the DFT of a signal $x(k)$ of length M.

$$\gamma_x^2 \simeq \frac{\exp\left(\frac{1}{M}\sum_{\mu=0}^{M-1} \ln |X(e^{j\Omega_\mu})|^2\right)}{\frac{1}{M}\sum_{\mu=0}^{M-1} |X(e^{j\Omega_\mu})|^2}$$

$$= \frac{\left[\prod_{\mu=0}^{M-1} |X(e^{j\Omega_\mu})|^2\right]^{\frac{1}{M}}}{\frac{1}{M}\sum_{\mu=0}^{M-1} |X(e^{j\Omega_\mu})|^2}, \quad \Omega_\mu = \frac{2\pi}{M}\mu, \quad \mu = 0, 1, \cdots, M-1.$$

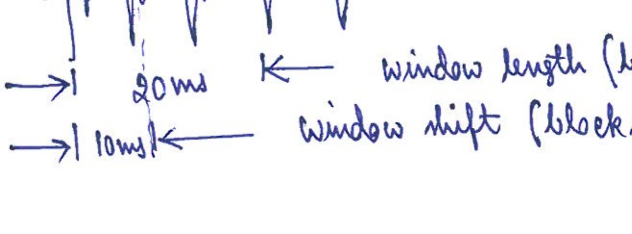
For a minimum phase system, with an LP-analysis filter with $A_0(z) = 1 - \cancel{A}(z) = 1 - \sum_{i=1}^{n} a_i z^{-i} = \frac{1}{z^n}\prod_{i=1}^{n}(z - z_i)$, all the zeros are in unit circle. In this case, we get a relation

$$\lim_{n \to \infty} \left\{\frac{1}{G_p}\right\} = \gamma_x^2$$

Predictor Adaptation

- Speech signals can be considered as stationary for a relatively short time intervals between 10 ms to 400 ms.
- The LP system parameters change frequently
- So, we need to update the parameters (mainly LPC coefficients)

Block-oriented Adaptation



window length (block length).

window shift (block shift).

We first define an window of length L

$$w(k) = \begin{cases} \text{non-zero value} & 0 \leq k \leq L-1 \\ 0 & k \geq L \text{ and } k < 0. \end{cases}$$

So, we have a new signal

$$\tilde{x}(k) = x(k) \cdot w(k-s) \qquad [S: \text{shift}]$$

Now, we recall $\phi_{xx}(\lambda) = \mathcal{E}\{x(k)\, x(k+\lambda)\}$

Note: We want to model the $\mathcal{E}\{\cdot\}$ by finite samples and we also have windowed sequence $\tilde{x}(k)$.

<u>Auto-correlation Method</u>

auto-correlation $r_\lambda = \sum_k \tilde{x}(k)\, \tilde{x}(k+\lambda)$

We also get a symmetry $r_\lambda = r_{-\lambda}$.

So, r_λ replaces $\phi_{xx}(\lambda)$ in Normal Equations and we solve LPCs.

$$\underbrace{\begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{n-1} \\ r_1 & r_0 & & \cdots & r_{n-2} \\ \vdots & & \ddots & & \vdots \\ & & & & r_0 \\ r_{n-1} & \cdots & & & \end{bmatrix}}_{\text{Symmetric and Toeplitz}} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

- For Large 'n', inversion is computationally expensive and also introduces numerical errors.
- There is a sophisticated algorithm: Levin-Durbin Recursion

<u>Covariance Method</u>

Let students read by themselves (Home work)

Sequential Adaptation

- For block-oriented adaptation, the LPCs $\{a_i\}_{i=1}^{n}$ are recalculated for each block or windowed portion.

- But we can adapt the LPCs for each sample time, given the full knowledge of LPCs in the previous sample time.

- $\boxed{\text{So what we want: an easy update of } a_i(k+1) \text{ using } a_i(k).}$

- Least-mean-square (LMS) algorithm is a nice approach.

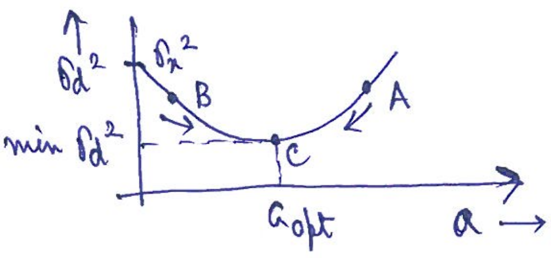
~~One step~~

first order predictor case

$$d(k) = x(k) - a\,x(k-1)$$

$$\sigma_d^2 = \sigma_x^2 - 2a\,\phi_{xx}(1) + a^2\,\phi_{xx}(0) \qquad [\text{Note } \phi_{xx}(0) = \sigma_x^2]$$

- σ_d^2 is a second order function of 'a'.



$$a_{opt} = \frac{\phi_{xx}(1)}{\phi_{xx}(0)}$$

Starting from points A or B, the minimum point (C) can be approached iteratively by taking the gradient

$$\nabla = \frac{\partial \sigma_d^2}{\partial a} = -2\,\phi_{xx}(1) + 2a\,\phi_{xx}(0).$$

$$= 2\,\phi_{xx}(0)\,.\,(a - a_{opt})$$

$$\nabla \propto (a - a_{opt})$$

- The instantaneous coefficient $a(k)$ must be corrected in the direction of the negative gradient according to

$$a(k+1) = a(k) - \eta . \nabla$$

Here η is a suitable step-size.

<u>What happens for general LP?</u>

$$\underline{a}(k) = (a_1(k), a_2(k), \cdots a_n(k))^T$$

$$\hat{x}(k) = \underline{a}^T(k) \; \underline{x}(k-1)$$

$$\sigma_d^2 = \mathcal{E}\{(x(k) - \underline{a}^T(k) \; \underline{x}(k-1))^2\}$$

$$\nabla = \frac{\partial \sigma_d^2}{\partial \underline{a}} = -2 \; \mathcal{E}\{(x(k) - \underline{a}^T(k) \; \underline{x}(k-1)) \; \underline{x}(k-1)\}$$

$$= -2\phi_{xx} + 2 R_{xx} \underline{a}(k)$$

So, again

$$\underline{a}(k+1) = \underline{a}(k) - \eta . \nabla$$

$$= \underline{a}(k) + 2\eta \; (\phi_{xx} - R_{xx} \underline{a}(k))$$

- The above approach uses expectation. So, in reality, knowledge of many signal samples are required.

<u>LMS:</u>

- instantaneous approach.

$$\hat{\sigma}_d^2(k) = d^2(k) = (x(k) - \underline{a}^T(k) \; \underline{x}(k-1))^2$$

$$\hat{\nabla} = +2\, d(k) \frac{\partial d(k)}{\partial \underline{a}(k)} = -2\, d(k) \; \underline{x}(k-1)$$

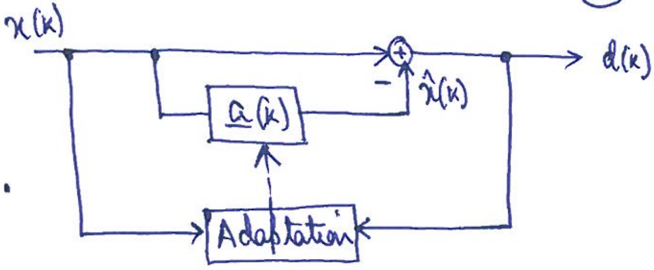
$$\boxed{\underline{a}(k+1) = \underline{a}(k) - \eta . \hat{\nabla} = \underline{a}(k) + 2\eta \; d(k) \; \underline{x}(k-1)}$$

For individual co-eff.

$$a_i(k+1) = a_i(k) + 2\eta \; d(k) \; x(k-i) \quad , i = 1, \cdots, n$$

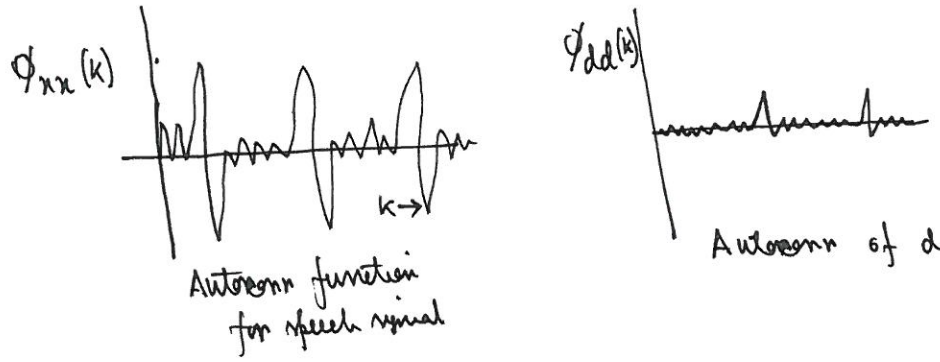
For stability reasons,

$$0 < \eta < \frac{1}{\|\underline{x}(k-1)\|_2^2}.$$

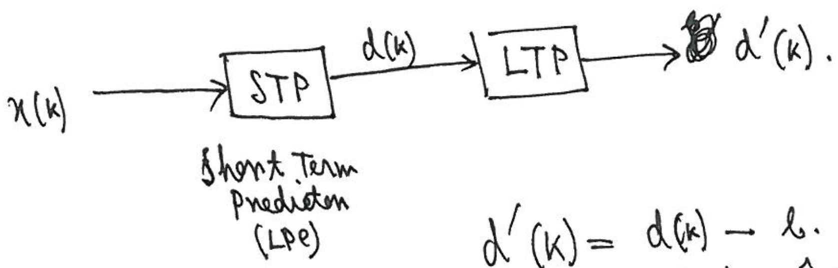


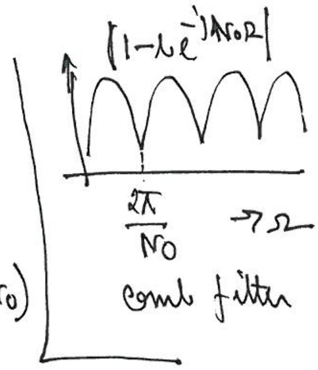
Long-Term Prediction (LTP)

- According to speech production model, the LP synthesis filter represents the vocal tract and the prediction error signal $d(k)$ is the excitation.
- for voiced speech, excitation signal is modeled as a pulse train.
- The prediction error signal exhibits long term correlation, which is due to period $T_0 = \frac{1}{f_0}$ for voiced segments



- With fundamental frequency $50\,Hz \leq f_0 \leq 250\,Hz$ and a sampling rate $f_s = 8\,KHz$, the periods T_0 have lengths between $N_0 = 32 - 160$ samples.
- The prediction over T_0 is called LTP.





$$d'(k) = d(k) - b \cdot d(k-N_0)$$
$$= d(k) - \hat{d}(k).$$

LTP filter Frequency response $\frac{d'(e^{j\Omega})}{d(e^{j\Omega})} = \underbrace{1 - b\, e^{-jN_0\Omega}}_{\text{A comb filter response.}}$

Q: what are the parameters in LTP?

Ans: b, N_0.

We have to choose them to minimize

$$\sigma_{d'}^2 = \sum_K d'^2(K) = \sum_K \left(d(K) - b\, d(K-N_0)\right)^2$$

$$\frac{\partial}{\partial b}\sigma_{d'}^2 = \sum_K \left(d(K) - b\, d(K-N_0)\right)\left(-d(K-N_0)\right) = 0$$

or, $$\sum_K d(K)\, d(K-N_0) = b\sum_K d(K-N_0)\, d(K-N_0)$$

$$\boxed{\text{or, } b = \frac{\sum_K d(K)\, d(K-N_0)}{\sum_K d^2(K-N_0)} = \frac{\text{ACR of } d(K) \text{ with } N_0 \text{ lag}}{\text{energy of } d(K)} = \frac{R(N_0)}{S(N_0)}}$$

Now, minimum $$\sigma_{d'}^2 = \sum_K d^2(K) - 2\sum_K d(K)\, b\, d(K-N_0) + \sum_K b^2 d^2(K-N_0)$$

$$= \sum_K d^2(K) - 2b\, R(N_0) + b^2\, S(N_0)$$

$$= S(0) - 2\frac{R(N_0)}{S(N_0)}\cdot R(N_0) + \frac{R^2(N_0)}{S^2(N_0)}\cdot S(N_0)$$

$$= S(0) - 2\frac{R^2(N_0)}{S(N_0)} + \frac{R^2(N_0)}{S(N_0)}$$

$$= S(0) - \frac{R^2(N_0)}{S(N_0)}.$$

So, $\boxed{\text{we should pick that "}N_0\text{" which minimizes } S(0) - \frac{R^2(N_0)}{S(N_0)}}$.