Linear Predictive Analysis - I

Recursive Formula for the Cepstrum

Taking the derivative of $\hat{X}(z) \triangleq \log X(z)$

$$\frac{d\hat{X}(z)}{dz} = \frac{d}{dz}\log X(z) = \frac{1}{X(z)}\frac{dX(z)}{dz} \quad \Rightarrow \quad \frac{dX(z)}{dz} = X(z)\frac{d\hat{X}(z)}{dz}, \tag{1}$$

and noting that the inverse z-transform of $-z\frac{dX(z)}{dz}$ is $nx[n]$ (and of $-z\frac{d\hat{X}(z)}{dz}$ is $n\hat{x}[n]$), we have

$$nx[n] = x[n] * (n\hat{x}[n]) = \sum_k k\hat{x}[k]x[n-k]. \tag{2}$$

Suppose $x[n] = 0$ and $\hat{x}[n] = 0$ for $n < 0$ (e.g. $x[n]$ is minimum-phase). For $n = 0$, using the initial value theorem of z-transforms, we have

$$\lim_{z\to\infty} \hat{X}(z) = \lim_{z\to\infty} \log X(z) \quad \Rightarrow \quad \hat{x}[0] = \log x[0]. \tag{3}$$

For $n > 0$, we have

$$\begin{aligned} nx[0]\hat{x}[n] &= nx[n] - \sum_{k=0}^{n-1} k\hat{x}[k]x[n-k] \\ \Rightarrow \quad \hat{x}[n] &= \frac{x[n]}{x[0]} - \sum_{k=0}^{n-1}\left(\frac{k}{n}\right)\frac{x[n-k]}{x[0]}\hat{x}[k] \end{aligned} \tag{4}$$

which is a recursive relation for $\hat{x}[n]$.

All-Pole Model

The all-pole model for the vocal tract system is of the form

$$H(z) = \frac{G}{A(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{G}{\prod_{k=1}^{p}(1 - z_k z^{-1})}, \tag{5}$$

where $z_k,\ k = 1, \ldots, p$ are all poles. Taking the inverse z-transform, the corresponding difference equation for (5) is

$$H(z)\left(1 - \sum_{k=1}^{p} a_k z^{-k}\right) = G \quad \Rightarrow \quad h[n] = \sum_{k=1}^{p} a_k h[n-k] + G\delta[n]. \tag{6}$$

Linear Predictive Analysis

Basic Idea
The linear predictive analysis (a.k.a. linear predictive coding, LPC) is based on the assumption that a speech sample can be approximated as a linear combination of p past samples.

$$s[n] \approx \tilde{s}[n] = \sum_{k=1}^{p} \alpha_k s[n-k]. \tag{7}$$

The linear predictive coefficients

$$\alpha_1, \ldots, \alpha_p \tag{8}$$

are determined by minimizing the squared error of the predicted signal and the actual signal.

Theoretical Ground
LPC is based on the simplified speech production model that we are familiar with now. Note that we have bundled the glottal pulse function $g[n]$, the vocal tract impulse response $v[n]$ and the radiation load response $r[n]$ together, which is denoted as $h[n]$. $H(z)$ is called the *vocal tract system function*. The excitation is denoted as $u[n]$, which is the impulse function or unit-variance random noise, depending on the voicing of speech.

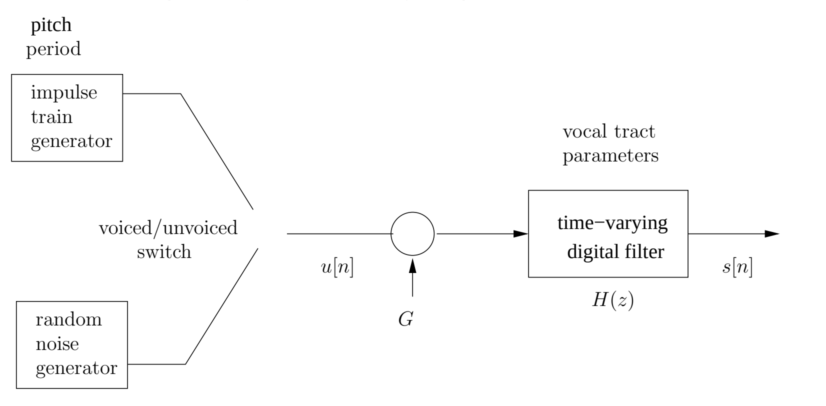

Figure: Simplied Speech Production Model

From the above figure, we have

$$S(z) = GU(z)H(z) \quad \Rightarrow \quad H(z) = \frac{S(z)}{GU(z)}. \tag{9}$$

In LPC, $H(z)$ is approximated by an all-pole function

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}, \tag{10}$$

so we have the time-domain LCCDE

$$s[n] = \sum_{k=1}^{p} a_k s[n-k] + Gu[n]. \tag{11}$$

Prediction Error
The prediction error is defined as

$$e[n] \triangleq s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^{p} \alpha_k s[n-k]. \tag{12}$$

Taking the z-transform, we have

$$E(z) = S(z) - \sum_{k=1}^{p} \alpha_k z^{-k} S(z) \quad \Rightarrow \quad \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \triangleq A(z). \tag{13}$$

$A(z)$ is called the *predictor error polynomial*.

Perfect Estimation
Perfect estimation means

$$\alpha_k = a_k, \quad k = 1, \ldots, p. \tag{14}$$

In this case, we have

$$e[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k] = Gu[n]. \tag{15}$$

Therefore, $E(z) = GU(z)$, and we have

$$A(z) = \frac{1}{H(z)}. \tag{16}$$

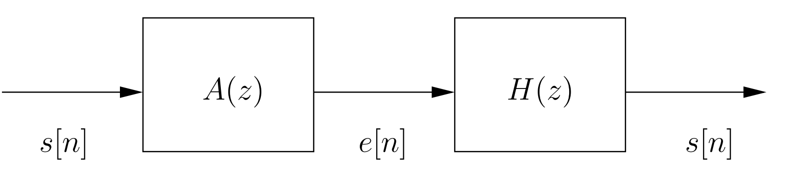

Figure: Inverse Filtering

Basic Formulation for LPC Equations
The *short-time* squared prediction error is defined as

$$\mathcal{E}_{\hat{n}}(\alpha_1, \ldots, \alpha_p) = \sum_m e_{\hat{n}}^2[m] = \sum_m \left(s_{\hat{n}}[m] - \sum_{k=1}^{p} \alpha_k s_{\hat{n}}[m-k]\right)^2, \tag{17}$$

where $\hat{n}$ is the index for short-time processing. To find the minimum $\mathcal{E}(\hat{\alpha}_1, \ldots, \hat{\alpha}_p)$, we take the derivative of (17) with respect to α_i and set it to 0 to get

$$\sum_m s_{\hat{n}}[m-i]s[m] = \sum_{k=1}^{p} \hat{\alpha}_k \sum_m s_{\hat{n}}[m-i]s_{\hat{n}}[m-k], \quad 1 \le i \le p. \tag{18}$$

Defining the auto-correlation function

$$\varphi_{\hat{n}}[i,k] = \sum_m s_{\hat{n}}[m-i]s_{\hat{n}}[m-k], \tag{19}$$

(18) can be re-written as

$$\sum_{k=1}^{p} \hat{\alpha}_k \varphi_{\hat{n}}[i,k] = \varphi_{\hat{n}}[i,0], \tag{20}$$

which is a problem in linear algebra! At this minimum, the squared prediction error is

$$\begin{aligned} \hat{\mathcal{E}}_{\hat{n}} &= \sum_m s_{\hat{n}}^2[m] - \sum_{k=1}^{p}\sum_m \hat{\alpha}_k s_{\hat{n}}[m]s_{\hat{n}}[m-k] \\ &= \varphi_{\hat{n}}[0,0] - \sum_{k=1}^{p} \hat{\alpha}_k \varphi_{\hat{n}}[0,k] \end{aligned} \tag{21}$$

Computation of LPC Coefficients

(20) is a system of linear equations in variables α_k and coefficients $\varphi_{\hat{n}}[i,k]$. The solution is quite straightforward once $\varphi_{\hat{n}}[i,k]$ is decided. There are several methods which differ mainly in how the error function (17) is actually computed.

The Autocorrelation Method
In this approach, *the value of signal is set to* 0 *outside the analysis window*,

$$s_{\hat{n}}[m] = s[m+\hat{n}]w[m], \quad 0 \le m \le L-1, \tag{22}$$

and the sum of the squared errors is over the entire signal, but is effectively $L+p$

$$\mathcal{E}_{\hat{n}} = \sum_{m=-\infty}^{\infty} e_{\hat{n}}^2[m] = \sum_{m=0}^{L-1+p} e_{\hat{n}}^2[m] \tag{23}$$

since $e_{\hat{n}}^2[m] = 0$ outside the range $[0, L-1+p]$. Applying the general results in the case, $\varphi_{\hat{n}}[i,k]$ in (19) becomes

$$\begin{aligned} \varphi_{\hat{n}}[i,k] &= \sum_{m=0}^{L-1+p} s_{\hat{n}}[m-i]s_{\hat{n}}[m-k] = \sum_{m=0}^{L-1-i+p} s_{\hat{n}}[m]s_{\hat{n}}[m+i-k] \\ &= \sum_{m=0}^{L-1-i+k} s_{\hat{n}}[m]s_{\hat{n}}[m+i-k], \quad 0 \le i,k \le p \end{aligned} \tag{24}$$

which is the short-time auto-correlation of $s_{\hat{n}}[m]$ evaluated for $(i-k)$. So we can write

$$\varphi_{\hat{n}}[i,k] = R_{\hat{n}}[i-k], \quad 0 \le i,k \le p. \tag{25}$$

Note that $R_{\hat{n}}[k]$, the short-time auto-correlation function is symmetric,

$$R_{\hat{n}}[i-k] = R_{\hat{n}}[|i-k|], \quad 0 \le i,k \le p. \tag{26}$$

It follows we have the following system of linear equations

$$\sum_{k=1}^{p} \hat{\alpha}_k R_{\hat{n}}[|i-k|] = R_{\hat{n}}[i], \quad 1 \le i \le p, \tag{27}$$

and the following minimum squared prediction error

$$\hat{\mathcal{E}}_{\hat{n}} = R_{\hat{n}}[0] - \sum_{k=1}^{p} \hat{\alpha}_k R_{\hat{n}}[k] \tag{28}$$

(27) can be written in a matrix form as follows

$$\begin{bmatrix} R_{\hat{n}}[0] & R_{\hat{n}}[1] & \ldots & R_{\hat{n}}[p-1] \\ R_{\hat{n}}[1] & R_{\hat{n}}[0] & \ldots & R_{\hat{n}}[p-2] \\ R_{\hat{n}}[2] & R_{\hat{n}}[1] & \ldots & R_{\hat{n}}[p-3] \\ \vdots & \vdots & \vdots & \vdots \\ R_{\hat{n}}[p-1] & R_{\hat{n}}[p-2] & \ldots & R_{\hat{n}}[0] \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} R_{\hat{n}}[1] \\ R_{\hat{n}}[2] \\ R_{\hat{n}}[3] \\ \vdots \\ R_{\hat{n}}[p] \end{bmatrix} \tag{29}$$

The Covariance Method
In this alternative approach, *the number of terms in the sum of the squared errors is fixed by the window size L.* The squared prediction error function is

$$\mathcal{E}_{\hat{n}}(\alpha_1, \ldots, \alpha_p) = \sum_{m=0}^{L-1} \left(s_{\hat{n}}[m] - \sum_{k=1}^{p} \alpha_k s_{\hat{n}}[m-k]\right)^2, \tag{30}$$

Applying the general results in the case, $\varphi_{\hat{n}}[i,k]$ becomes

$$\varphi_{\hat{n}}[i,k] = \phi_{\hat{n}}[i,k] = \sum_{m=0}^{L-1} s_{\hat{n}}[m-i]s_{\hat{n}}[m-k], \quad 0 \le i,k \le p \tag{31}$$

Noting that $s_{\hat{n}}[m] = s[\hat{n}+m]$, we have

$$\phi_{\hat{n}}[i,k] = \sum_{m=-i}^{L-i-1} s_{\hat{n}}[m]s_{\hat{n}}[m+i-k] = \sum_{m=-k}^{L-k-1} s_{\hat{n}}[m]s_{\hat{n}}[m+k-i] \tag{32}$$

It follows we have the following system of linear equations

$$\sum_{k=1}^{p} \hat{\alpha}_k \phi_{\hat{n}}[i,k] = \phi_{\hat{n}}[i,0], \quad 1 \le i \le p, \tag{33}$$

and the following minimum squared prediction error

$$\hat{\mathcal{E}}_{\hat{n}} = \phi_{\hat{n}}[0,0] - \sum_{k=1}^{p} \hat{\alpha}_k \phi_{\hat{n}}[0,k] \tag{34}$$

(33) can be written in a matrix form as follows

$$\begin{bmatrix} \phi_{\hat{n}}[1,1] & \phi_{\hat{n}}[1,2] & \ldots & \phi_{\hat{n}}[1,p] \\ \phi_{\hat{n}}[2,1] & \phi_{\hat{n}}[2,2] & \ldots & \phi_{\hat{n}}[2,p] \\ \phi_{\hat{n}}[3,1] & \phi_{\hat{n}}[3,2] & \ldots & \phi_{\hat{n}}[3,p] \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{\hat{n}}[p,p] & \phi_{\hat{n}}[p,p] & \ldots & \phi_{\hat{n}}[p,p] \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \phi_{\hat{n}}[1,0] \\ \phi_{\hat{n}}[2,0] \\ \phi_{\hat{n}}[3,0] \\ \vdots \\ \phi_{\hat{n}}[p,0] \end{bmatrix} \tag{35}$$

Note that the matrix in (29) is Toeplitz, while in (35) is symmetric and positive definite. The latter property is also possessed by a covariance matrix, hence the name *covariance method*.

Computation of LPC Gain

Suppose $\tilde{H}(z)$ is the system function with

$$\tilde{H}(z) = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}} \tag{36}$$

where $\alpha_k,\ k = 1, \ldots, p$ are estimated, say by the autocorrelation method on a speech window. The impulse response of such a system satisfies the difference equation

$$\tilde{h}[n] = \sum_{k=1}^{p} \alpha_k \tilde{h}[n-k] + G\delta[n]. \tag{37}$$

Denote the autocorrelation function of $\tilde{h}[n]$ as

$$\tilde{R}[m] = \sum_n \tilde{h}[n]\tilde{h}[n+m] = \tilde{R}[-m]. \tag{38}$$

It can be shown that $\tilde{R}[m]$ satisfies the linear equations as the short-time autocorrelation function of the speech signal, (27), and thus is proportional and $R_{\hat{n}}[m]$. Furthermore

$$\tilde{R}[0] = \sum_{k=1}^{p} \alpha_k \tilde{R}[k] + G^2 \tag{39}$$

Autocorrelation Matching Property
This is a condition that

$$\tilde{R}[m] = R_{\hat{n}}[m]. \tag{40}$$

It follows that the gain is

$$G^2 = \tilde{R}[0] - \sum_{k=1}^{p} \alpha_k \tilde{R}[k] = R_{\hat{n}}[0] - \sum_{k=1}^{p} \alpha_k R_{\hat{n}}[k] = \hat{\mathcal{E}}_{\hat{n}}. \tag{41}$$

Frequency Domain Interpretation

The frequency response of the vocal tract system function is approximated by

$$\tilde{H}(e^{j\omega}) = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k e^{-j\omega k}}. \tag{42}$$

Due to the fact that $\tilde{R}[m]$ and $R_{\hat{n}}[m]$ are identical for $m = 1, \ldots, p$, we have

$$\lim_{p\to\infty} |\tilde{H}(e^{j\omega})|^2 = |S_{\hat{n}}(e^{j\omega})|^2, \tag{43}$$

so $|\tilde{H}(e^{j\omega})|$ is an approximation to $|S_{\hat{n}}(e^{j\omega})|$, the STFT of the window at $\hat{n}$,

$$S_{\hat{n}}(e^{j\omega}) = \sum_m s[\hat{n}+m]w[m]e^{-j\omega m}. \tag{44}$$

Essentially, $|\tilde{H}(e^{j\omega})|$ is a "smoothed" version of $S_{\hat{n}}(e^{j\omega})|$, with the excitation structure removed. Figure 9.5 and Figure 9.6 show voiced and unvoiced examples.

Squared Prediction Error in the Frequency Domain
Recall that in autocorrelation method, the squared prediction error is

$$\mathcal{E}_{\hat{n}} = \sum_{m=0}^{L-1+p} e_{\hat{n}}^2[m]. \tag{45}$$

(45) translates into

$$\mathcal{E}_{\hat{n}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} |E_{\hat{n}}(e^{j\omega})|^2 d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{|S_{\hat{n}}(e^{j\omega})|^2}{|\tilde{H}(e^{j\omega})|^2} d\omega \tag{46}$$

in the frequency domain. According to (46), the minimization of $\mathcal{E}_{\hat{n}}$ to determine tends $\tilde{H}(e^{j\omega})$ to emphasize the high speech energy part and de-emphasize the low speech energy part. Figure 9.7.