ECE 537 Fundamentals of Speech Processing
Problem Set 5

UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering

Assigned: Monday, 10/3/2022; Due: Monday, 10/10/2022
Reading: Atal, "Predictive Coding of Speech at Low Bit Rates," 1982

1. Equation (4) in the article gives a formula for the prediction sum-squared error (a.k.a. the energy of the prediction residual) of an $(m-1)$-tap linear prediction. To see why this is the case, let's explore the origin of some of the equations in this section.

   (a) (1 point) Define ϵ_m in the following way:

   $$\epsilon_m = \sum_{n=p+1}^{N+p} (d_n^{(m-1)})^2,$$

   where

   $$d_n^{(m)} = s_n - \sum_{k=1}^{m-1} a_k^{(m-1)} s_{n-k},$$

   and $a_k^{(m)}$ is the k^{th} coefficient of an order-m linear predictor. Solve for $\frac{\partial \epsilon_m}{\partial a_k^{(m-1)}}$ in terms of the coefficients $a_i^{(m)}$ and the speech signal s_n.

   Solution:

   $$\frac{\partial \epsilon_m}{\partial a_k^{(m-1)}} = -2 \sum_{n=p+1}^{N+p} s_n s_{n-k} + 2 \sum_{n=p+1}^{N+p} \sum_{i=1}^{m-1} a_i^{(m-1)} s_{n-i} s_{n-k}$$

   (b) (1 point) Consider setting $\frac{\partial \epsilon_m}{\partial a_k^{(m-1)}} = 0$ simultaneously for all $k \in \{1, \dots, m\}$; this results in m linear equations in terms of s_n and $a_i^{(m-1)}$. Convert these m linear equations into a single matrix equation in terms of the vector $\vec{a}_{m-1} = [a_1^{(m-1)}, \dots, a_m^{(m-1)}]^T$, the vector $\vec{c}_{m-1} = [\phi(0,1), \dots, \phi(0, m-1)]^T$, and the matrix Φ_{m-1} defined as

   $$\Phi_{m-1} = \begin{bmatrix} \phi(1,1) & \cdots & \phi(1, m-1) \\ \vdots & \ddots & \vdots \\ \phi(m-1, 1) & \cdots & \phi(m-1, m-1) \end{bmatrix},$$

   where

   $$\phi(i,j) = \sum_{n=p+1}^{N+p} s_{n-i} s_{n-j}$$

Solution:
$$\Phi_{m-1}\vec{a}_{m-1} = \vec{c}_{m-1}$$

(c) (1 point) Notice that the equation $\frac{\partial \epsilon_m}{\partial a_k^{(m-1)}} = 0$ can be written as

$$\sum_{n=p+1}^{N+p} d_n^{(m-1)} s_{n-k} = 0 \qquad (1)$$

Eq. (1) is called the orthogonality condition. It says that the coefficient $a_k^{(m-1)}$ that minimizes ϵ_m is the one that eliminates all correlation between $d_n^{(m-1)}$ (the prediction residual) and s_{n-k} (the predictor). Use the orthogonality condition to write ϵ_m as an affine function of the coefficients $a_i^{(m-1)}$, where the coefficients in the linear function are the covariance terms $\phi(i,j)$. If your equation has any terms that are quadratic in $a_i^{(m-1)}$, then you haven't simplified it far enough, keep going.

Solution:

$$\begin{aligned}
\epsilon_m &= \sum_{n=p+1}^{N+p} (d_n^{(m-1)})^2 \\
&= \sum_{n=p+1}^{N+p} d_n^{(m)} \left(s_n - \sum_{k=1}^{m-1} a_k^{(m-1)} s_{n-k} \right) \\
&= \sum_{n=p+1}^{N+p} d_n^{(m-1)} s_n \\
&= \sum_{n=p+1}^{N+p} \left(s_n - \sum_{k=1}^{m-1} a_k^{(m-1)} s_{n-k} \right) s_n \\
&= \phi(0,0) - \sum_{k=1}^{m-1} a_k^{(m-1)} \phi(0,k)
\end{aligned}$$

(d) (1 point) The covariance LPC method solves the equation $\Phi\vec{a} = \vec{c}$ in three steps. First, it computes the Cholesky decomposition $\Phi = LL^T$, where L is a lower-triangular matrix. The Cholesky decomposition is an $\mathcal{O}\{p^3\}$ operation. Second, it solves for the vector $\vec{q}$ in the equation $L\vec{q} = \vec{c}$; this is an $\mathcal{O}\{p^2\}$ operation. Third, it solves for $\vec{a}$ either directly, by solving the equation $\vec{q} = L^T\vec{a}$, or indirectly, using the "known relation between the partial correlations and the predictor coefficients" that is named but not described in the article; in either case, this is an $\mathcal{O}\{p^2\}$ operation.

The part of all this that's interesting to us is that, if we define the order-m equations as $\Phi_{m-1} = L_{m-1}L_{m-1}^T$, $L_{m-1}\vec{q}_{m-1} = \vec{c}_{m-1}$, and $\vec{q}_{m-1} = L_{m-1}^T\vec{a}_{m-1}$, then we get that

$$\sum_{i=1}^{m-1} q_i^2 = \vec{q}_{m-1}^T\vec{q}_{m-1} = \vec{a}_{m-1}^T\Phi_{m-1}\vec{a}_{m-1} = \vec{a}_{m-1}^T\vec{c}_{m-1} \qquad (2)$$

Use Eq. (2) to re-write your solution to part (c) in terms of the coefficients q_i.

Digression: here is an observation that is not necessary to solve this problem, but that might help you to deepen your understanding of LPC. In the Cholesky decomposition $\Phi = LL^T$, the m^{th} row of L only depends on the first m rows and columns of Φ, therefore L_{m-1} is a submatrix of L_m. Similarly, in the equation $L\vec{q} = \vec{c}$, the m^{th} coefficient, q_m, depends only on the first m rows of L, and the first m elements of $\vec{c}$, therefore the vector $\vec{q}_{m-1}$ is a subvector of $\vec{q}_m$. The same is not true of the equation $L^T\vec{q}_m = \vec{a}_m$. The i^{th} predictor coefficient of an order-m predictor, is therefore **not the same** as the i^{th} coefficient of an order-p predictor:

$$a_i^{(m)} \neq a_i^{(p)},$$

but the partial correlation coefficient q_i/ϵ_i is the same for any order, m or p, as long as $m \geq i$ and $p \geq i$.

Solution:

$$\sum_{i=1}^{m-1} q_i^2 = \vec{a}_{m-1}^T \vec{c}_{m-1} = \sum_{k=1}^{m-1} a_k^{(m-1)} \phi(0,k)$$

Therefore

$$\epsilon_m = \phi(0,0) - \sum_{i=1}^{m-1} q_i^2$$

2. (1 point) Suppose that d_n is the LPC residual, and v_n is the pitch prediction residual, thus

$$v_n = d_n - \beta_1 d_{n-M+1} - \beta_2 d_{n-M} - \beta_3 d_{n-M-1} \quad (3)$$

If d_n were perfectly periodic with a period of M, then it would be possible to set $v_n = 0$ (after the first pitch period) by simply choosing $\beta_1 = 0, \beta_2 = 1, \beta_3 = 0$. The reason Eq. (3) contains three delay terms, instead of just one, is that the pitch period might not be an integer.

Suppose that d_n is perfectly periodic, but with a period τ that is not an integer. In this case, the ideal pitch predictor should be $P_d(e^{j\omega}) = e^{-j\omega\tau}$ in the range $-\pi < \omega < \pi$. What is the inverse transform, $p_d[n]$, of this pitch predictor? What would be reasonable values to choose for M, β_1, β_2, and β_3?

Solution:

$$\begin{aligned} p_d[n] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega(n-\tau)} d\omega \\ &= \text{sinc}\left(\pi(n-\tau)\right) \end{aligned}$$

A reasonable choice for M would be the integer nearest to τ, and the coefficients could be set to

$$\begin{aligned} \beta_1 &= \text{sinc}\left(\pi(M-1-\tau)\right) \\ \beta_2 &= \text{sinc}\left(\pi(M-\tau)\right) \\ \beta_3 &= \text{sinc}\left(\pi(M+1-\tau)\right) \end{aligned}$$

3. The idea of perceptual noise shaping is to shape the speech signal, producing $Y(\omega) = (1 - R(\omega))S(\omega)$, prior to quantizing it. Quantization generates a synthetic output, $\hat{q}_n$, in order to minimize

$$\epsilon = \sum_{n=p+1}^{N+p} q_n^2 = \sum_{n=p+1}^{N+p} (y_n - \hat{y}_n)^2,$$

where

$$\begin{aligned} \hat{Y}(\omega) &= \frac{1}{1 - P_A(\omega)} \hat{Q}(\omega) \\ \hat{S}(\omega) &= \frac{1}{1 - R(\omega)} \hat{Y}(\omega) \end{aligned}$$

(a) (1 point) Use Parseval's theorem to express ϵ as an integral, over frequency, of some function of $S(\omega)$, $\hat{S}(\omega)$, and $R(\omega)$.

Solution: By Parseval's theorem,

$$\sum_{n=p+1}^{N+p}(y_n-\hat{y}_n)^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|Y(\omega)-\hat{Y}(\omega)\right|^2 d\omega$$
$$= \frac{1}{\pi}\int_{0}^{\pi}\left|S(\omega)-\hat{S}(\omega)\right|^2 |1-R(\omega)|^2 d\omega$$

(b) (1 point) Usually, the quantization noise added in one time step, $q_n = \hat{y}_n - y_n$, is uncorrelated with the quantization noise added in any other time step, thus q_n is white noise, with some average power σ_q^2. Under the assumption that q_n is white noise with power σ_q^2, what is the power spectrum of $\hat{s}_n - s_n$?

Solution: If we can assume that

$$E\left[\left|\hat{Y}(\omega)-Y(\omega)\right|^2\right] = \sigma_q^2,$$

then

$$E\left[\left|\hat{S}(\omega)-S(\omega)\right|^2\right] = \frac{\sigma_q^2}{|1-R(\omega)|^2}.$$

(c) (1 point) For noise shaping, a reasonable set of principles might include:

1. Near a spectral pole of $S(\omega)$, it's OK to have louder noise, because the noise will be masked by the high energy of $S(\omega)$, thus the perceptual weighting $|1-R(\omega)|^2$ can be smaller at these frequencies, perhaps something like

$$1 \geq |1-R(\omega)|^2 \geq \frac{1}{|S(\omega)|^2} \text{ if } \omega \approx \omega_k,$$

   where ω_k is one of the spectral peaks of $S(\omega)$.

2. The perceptual weighting should be constant at frequencies far from any spectral pole, thus

$$|1-R(\omega)|^2 \approx 1 \text{ if } |\omega-\omega_k| \text{ is large}$$

Principle #2 is satisfied if $1-R(\omega)$ is an all-pass filter, i.e., its zeros have the same frequencies as its poles. Principle #1 is satisfied if the zeros and poles both have the frequencies of the LPC predictor, $1-P_A(z) = 1-\sum_{k=1}^{p} a_k z^{-k}$, and if the bandwidths of the poles are larger than the bandwidths of the zeros. To see that this is the case, consider the all-pass filter

$$1-R(z) = \frac{1-p_1 z^{-1}}{1-ap_1 z^{-1}}, \qquad (4)$$

where $p_1 = e^{-\sigma_1+j\omega_1}$, $\sigma_1 > 0$ is the (real-valued) half-bandwidth of the pole (measured in radians/sample), and a is some real constant in the range $0 \leq a < 1$. Show that $|1-R(e^{j\omega_1})| < 1$, where $|1-R(e^{j\omega_1})| < 1$ is the magnitude response of the all-pass filter at the frequency $\omega = \omega_1$.

Solution:

$$\begin{aligned} |1 - R(e^{j\omega_1})| &= \left| \frac{1 - p_1 e^{-j\omega_1}}{1 - ap_1 e^{-j\omega_1}} \right| \\ &= \left| \frac{1 - e^{-\sigma_1}}{1 - ae^{-\sigma_1}} \right| \end{aligned}$$

Since $0 < a < 1$, $0 < ae^{-\sigma_1} < e^{-\sigma_1} < 1$, $(1 - ae^{-\sigma_1}) > (1 - e^{-\sigma_1})$, and therefore $|1 - R(e^{j\omega_1})| < 1$.

(d) (1 point) Suppose that a speech signal has pole frequencies that are measured in radians/sample as ω_k, for $1 \leq k \leq p$, and corresponding bandwidths of $2\sigma_k$. (Assume that these are arranged in complex conjugate pairs, e.g., $\omega_{p+1-k} = -\omega_k$, and $\sigma_{p+1-k} = \sigma_k$). The LPC polynomial is therefore

$$1 - P_A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} = \prod_{i=1}^{p} (1 - p_i z^{-1}),$$

where $p_i = e^{-\sigma_i + j\omega_i}$. Eq. (21) in the article suggest using a perceptual weighting filter that has $1 - P_A(z)$ in the numerator, and the following denominator:

$$1 - \sum_{k=1}^{p} \alpha^k a_k z^{-k}, \tag{5}$$

where $0 \leq \alpha \leq 1$. Show that Eq. (5 has roots that have the same frequencies (ω_i) as $1 - P_A(z)$, and that its bandwidths have been increased to $\sigma_i - \ln \alpha$.

Solution:

$$1 - \sum_{k=1}^{p} \alpha^k a_k z^{-k} = \prod_{i=1}^{p} (1 - \alpha p_i z^{-1}),$$

The roots of this polynomial are

$$\alpha p_i = e^{\ln \alpha - \sigma_i + j\omega_i},$$

which has a center frequency of σ_i, and a bandwidth of $2(\sigma_i - \ln \alpha)$.