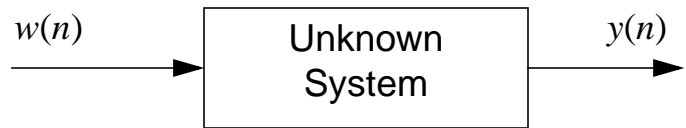


## The Orthogonality Principle and System Identification

Consider the problem of modeling the output of a system using an all-pole filter and a white noise input:



Let's assume that the system is of the form of an all-pole model:

$$y(n) = \sum_{k=1}^p \alpha_k y(n-k) + w(n)$$

If the model is sufficiently accurate,  $w(n)$  can be assumed to be white Gaussian noise. Suppose we are given measurements of the output,  $\tilde{y}(n)$ . Can we compute the coefficients,  $\{\alpha_k\}$ , that minimizes the error between our measurements and the linear prediction estimate of the measurements?

$$E = \sum_n (\tilde{y}(n) - y(n))^2 = \sum_n \left( \tilde{y}(n) - \left( \sum_{k=1}^p \alpha_k \tilde{y}(n-k) + w(n) \right) \right)^2$$

We can simplify the above expression by noting that:

$$\begin{aligned} E &= \sum_n (\tilde{y}(n))^2 - \sum_n \tilde{y}(n) \left( \sum_{k=1}^p \alpha_k \tilde{y}(n-k) \right) + \sum_n \tilde{y}(n) w(n) \\ &\quad - \sum_n \tilde{y}(n) \left( \sum_{k=1}^p \alpha_k \tilde{y}(n-k) \right) + \sum_n \left( \sum_{k=1}^p \alpha_k \tilde{y}(n-k) \right)^2 - \sum_n w(n) \sum_{k=1}^p \alpha_k \tilde{y}(n-k) \\ &\quad + \sum_n w(n) \tilde{y}(n) - \sum_n w(n) \sum_{k=1}^p \alpha_k \tilde{y}(n-k) + \sum_n w^2(n) \end{aligned}$$

Rearranging terms gives:

$$\begin{aligned} E &= r_{\tilde{y}\tilde{y}}(0) + r_{ww}(0) - 2 \sum_{k=1}^p \alpha_k r(k) + \sum_n \left( \sum_{k=1}^p \alpha_k \tilde{y}(n-k) \right)^2 \\ &\quad + 2r_{\tilde{y}w}(0) - 2 \left( \sum_{k=1}^p \alpha_k r_{\tilde{y}w}(k) \right) \end{aligned}$$

But, recall that the input to the system is white Gaussian noise, and hence is uncorrelated with the output signal. This implies all correlation terms between the signal and noise are zero. This argument is known as the orthogonality principle (the noise is orthogonal to the signal when the mean square-error predictor is used). Our expression for the error energy simplifies to:

$$E = r_{\tilde{y}\tilde{y}}(0) + r_{ww}(0) - 2 \sum_{k=1}^p \alpha_k r(k) + \sum_n \left( \sum_{k=1}^p \alpha_k \tilde{y}(n-k) \right)^2$$

Differentiating w.r.t to  $\alpha_l$  as before gives the normal linear prediction equation:

$$\bar{\alpha} = \underline{R}^{-1} \bar{r}$$

This view of the linear prediction problem is known as system identification (we are discovering an underlying model of the system using only measurements of the output). We can show that the LP estimate is, on the average, the estimator with the least error (or bias).

What if the system is not an all-pole system? We can approximate zeroes:

$$H(z) = 1 - az^{-1} \approx \frac{1}{1 + \alpha z^{-1} + (\alpha z^{-1})^2 + \dots}$$

It is also possible to postulate an autoregressive moving average model:

$$H(z) = \frac{1 + b_1 z^{-1} + \dots + b_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}}$$

Computation of these coefficients is a non-linear optimization problem that is difficult to solve efficiently and reliably. Marginal gains, if any, have been reported.

How does the linear prediction problem relate to more classical applications such as channel equalization (modems)?

What modifications do we need to make to the LP approach?

## Levinson-Durbin Recursion

The predictor coefficients can be efficiently computed for the autocorrelation method using the Levinson-Durbin recursion:

for  $i = 1, 2, \dots, p$

$$E_0 = r(0)$$

$$k_i = \left( r(i) - \sum_{j=1}^{i-1} a_{i-1}(j)r(i-j) \right) / E_{i-1}$$

for  $j = 1, 2, \dots, i-1$

$$a_i(i) = k_i$$

$$a_i(j) = a_{i-1}(j) - k_i a_{i-1}(i-j)$$

$$E_i = (1 - k_i^2) E_{i-1}$$

This recursion gives us great insight into the linear prediction process. First, we note that the intermediate variables,  $\{k_i\}$ , are referred to as *reflection coefficients*.

Example:  $p=2$

$$E_0 = r(0)$$

$$k_1 = r(1)/r(0)$$

$$a_1(1) = k_1 = r(1)/r(0)$$

$$\begin{aligned} E_1 &= (1 - k_1^2) E_0 \\ &= \frac{r^2(0) - r^2(1)}{r(0)} \end{aligned}$$

$$k_2 = \frac{r(2)r(0) - r^2(1)}{r^2(0) - r^2(1)}$$

$$a_2(2) = k_2 = \frac{r(2)r(0) - r^2(1)}{r^2(0) - r^2(1)}$$

$$\begin{aligned} a_2(1) &= a_1(1) - k_2 a_1(1) \\ &= \frac{r(1)r(0) - r(1)r(2)}{r^2(0) - r^2(1)} \end{aligned}$$

$$\alpha_1 = a_2(1)$$

$$\alpha_2 = a_2(2)$$

This reduces the LP problem to  $O(p)$  and saves an order of magnitude in computational complexity, and makes this analysis amenable to fixed-point digital signal processors and microprocessors.

## Linear Prediction Coefficient Transformations

The predictor coefficients and reflection coefficients can be transformed back and forth with no loss of information:

### Predictor to reflection coefficient transformation:

for  $i = p, p-1, \dots, 1$

$$k_i = a_i(i)$$

$$a_{i-1}(j) = \frac{a_i(j) + k_i a_i(i-j)}{1 - k_i^2} \quad 1 \leq j \leq i-1$$

### Reflection to predictor coefficient transformation:

for  $i = 1, 2, \dots, p$

$$a_i(i) = k_i$$

$$a_i(j) = a_{i-1}(j) - k_i a_{i-1}(i-j) \quad 1 \leq j \leq i-1$$

Also, note that these recursions require intermediate storage for  $\{a_i\}$ .

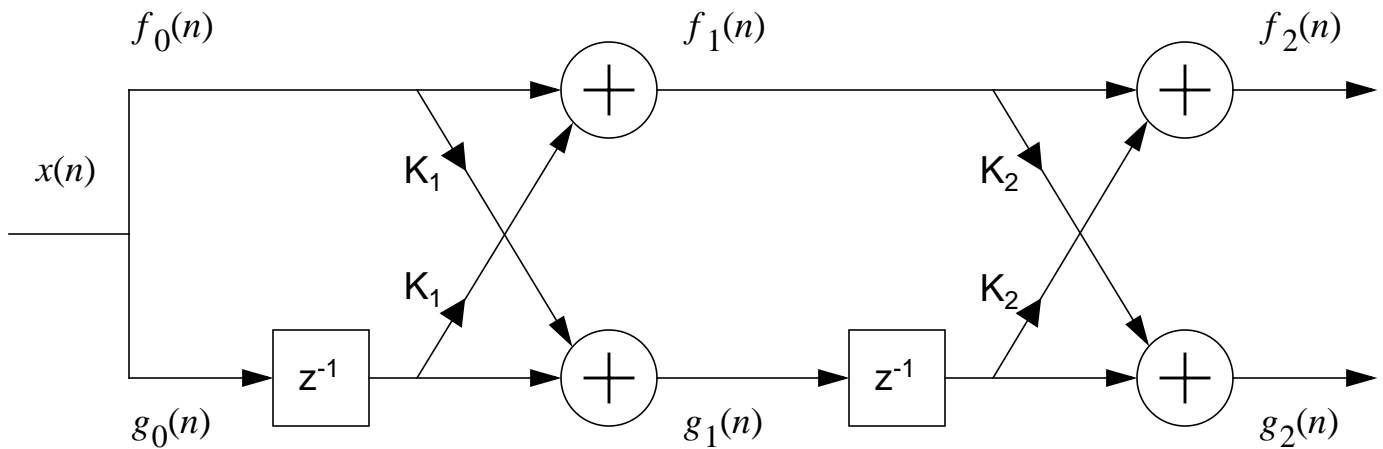
From the above recursions, it is clear that  $|k_i| \neq 1$ . In fact, there are several important results related to  $k_i$ :

- (1)  $|k_i| < 1$
- (2)  $|k_i| = 1$ , implies a harmonic process (poles on the unit circle).
- (3)  $|k_i| > 1$  implies an unstable synthesis filter (poles outside the unit circle).
- (4)  $E_0 > E_1 > \dots > E_p$

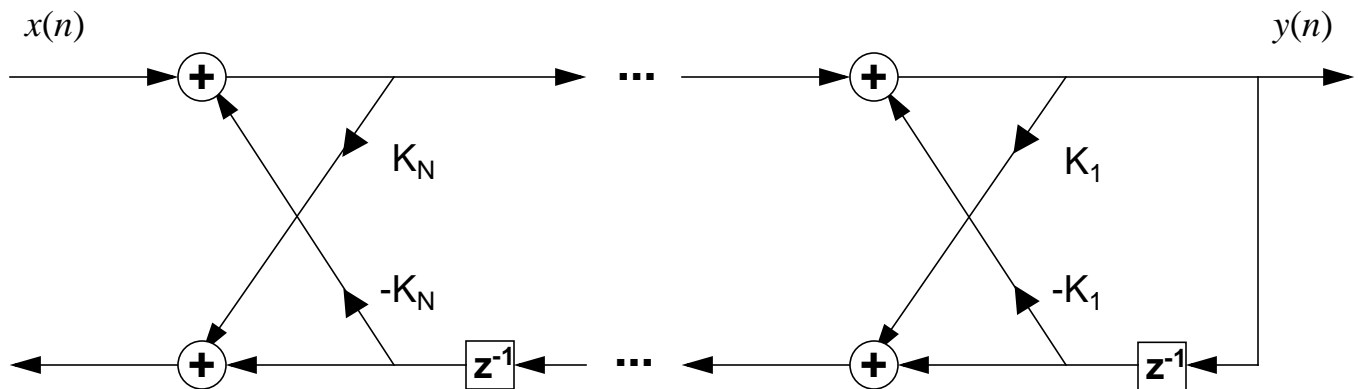
This gives us insight into how to determine the LP order during the calculations. We also see that reflection coefficients are orthogonal in the sense that the best order "p" model is also the first "p" coefficients in the order "p+1" LP model (very important!).

### Relationship to the Lattice Filters and Reflection Coefficients

The standard direct-form FIR filter can be implemented in a lattice structure:



The inverse, or Infinite Impulse Response (IIR) equivalent, is an all-pole filter:



The coefficients  $\{K_i\}$  are called reflection coefficients, and can be computed directly from the signal:

$$K_i = \frac{\sum_{m=0}^{N-1} f_{i-1}(m)g_{i-1}(m-1)}{\left\{ \left( \sum_{m=0}^{N-1} (f_{i-1}(m))^2 \right) \left( \sum_{m=0}^{N-1} (f_{i-1}(m-1))^2 \right) \right\}^{1/2}}$$

For the filter to be stable, these reflection coefficients must be bounded:  $|K_i| \leq 1$ .



## Generalized Lattice Solutions

One of the most famous lattice formulations is the Burg algorithm, originally introduced in the mid-60's prior to the introduction of LP in speech. There are actually a family of lattice solutions of a similar form.

The Burg algorithm requires the coefficients to be computed using:

$$K_i = \frac{\sum_{m=0}^{N-1} f_{i-1}(m)g_{i-1}(m-1)}{\frac{1}{2} \left\{ \sum_{m=0}^{N-1} f_{i-1}(m)^2 + \sum_{m=0}^{N-1} f_{i-1}(m-1)^2 \right\}}$$

We see this is a weighted average of the forward and backward error terms, and that the reflection coefficients are FORCED to be bounded.