

Lecture 12
Linear Prediction and All-Pole Modeling

As a prelude to the lecture, consider the following question:

A continuous-time sinusoidal signal which you want to sample is known to have frequency less than 5kHz. What's the minimum sampling frequency that will let you uniquely recover the signal from the samples?

The fact that 0 Hz is the answer often comes as a surprise to even seasoned DSP practitioners. Astonishment may be remedied by remembering that a sinusoid can be uniquely described by three parameters: frequency, amplitude, and phase. Taking three samples at $t = 0.1$ ms, $t = 0.2$ ms, and $t = 0.3$ ms therefore gives three equations with three unknowns, and the unknown parameters may be uniquely found. Since these three samples are the only ones taken for all time, the sampling frequency is 0 Hz.

The general business of describing a signal using reduced terms (and hopefully doing so well) is the thrust of the next two lectures. We'll formally refer to this practice as *parametric signal modeling*. The basic process of signal modeling is summed up in three steps:

- (1) Choose a model and associated parameters.
- (2) Define a modeling error.
- (3) Find a parameter set which minimizes the modeling error.

Remarkably, these three steps characterize a wide class of useful techniques which find application in many fields, including filter design; data compression and storage; signal identification, detection, and recognition; deconvolution; spectral analysis; and prediction.

The particular model which we will adopt over the course of this and the next lecture is concerned with representing an arbitrary signal $s[n]$ by a rational z -transform of the form

$$\boxed{S'(z) = \frac{A}{1 - \sum_{k=1}^P a_k z^{-k}}}, \quad (1)$$

where P is finite. Our associated model parameters are the a_k . This approximation to $s[n]$ is often referred to as an *all-pole model*. Of course there are many other ways to model a signal, such as representing it as a sinusoid, or even as a rational z -transform of the form

$$\hat{S}'(z) = \frac{A \sum_{k=0}^M b_k z^{-k}}{1 - \sum_{k=1}^N a_k z^{-k}}.$$

For a number of reasons though, some of which we'll get into, all-pole modeling has enjoyed success in practice.

Now that a model and associated parameters have been chosen, a modeling error must be defined. For a signal $s[n]$ and its corresponding all-pole model $s'[n]$, a first logical choice might be to define our modeling error $\hat{e}[n]$ as

$$\hat{e}[n] \equiv s[n] - s'[n],$$

where the a_k 's in Equation 1 are chosen to minimize

$$\hat{\varepsilon} = \sum_{n=-\infty}^{+\infty} \hat{e}^2[n].$$

As it turns out, optimizing $e[n]$ over the a_k 's in this case requires the solution of a set of nonlinear equations. It would therefore be convenient to either happen upon an elegant closed-form solution to these equations or to not have to solve them. Recalling that we have the luxury to define whatever modeling error we like, it makes sense to take on the problem of choosing a more mathematically convenient error metric.

Consider the scenario where the all-pole model $s'[n]$ perfectly models $s[n]$, *i.e.* where $s[n] = s'[n]$. How can we determine, given $s[n]$ and $s'[n]$, that $s[n]$ is perfectly modeled? One method involves first building the following system:

$$s[n] \longrightarrow \boxed{\frac{1}{A} [1 - \sum_{k=1}^p a_k z^{-k}]} \longrightarrow g[n].$$

If $s'[n]$ perfectly models $s[n]$, the output of this system must be $g[n] = \delta[n]$. We therefore define our modeling error as

$$\boxed{e[n] \equiv g[n] - \delta[n]}, \tag{2}$$

where

$$G(z) = \frac{S(z)}{S'(z)}.$$

How can we interpret this modeling error? To begin, note that we'd like $g[n]$ to be very close to 0 for $n \neq 0$. If this is the case, we know that $g[n] \approx C\delta[n]$, and so $s[n]$ is well-modeled by

$$s[n] = \sum_{k=1}^p a_k s[n-k], \quad n \neq 0.$$

If $s[n]$ fits the model, in other words, it is said to be *linearly predictable* from $s[n-1], s[n-2], \dots, s[n-p]$ for $n \neq 0$. For $n \neq 0$, $e[n]$ can therefore be thought of as a prediction error. Although there are some differences in the details, the problems of all-pole modeling and linear prediction lead to the same results.

As it turns out, finding the variables a_k such that

$$\varepsilon = \sum_{n=-\infty}^{+\infty} e^2[n]$$

is minimized requires the solution of a set of linear equations. We'll now prove this by solving the posed optimization problem.

The error metric

$$\varepsilon = \sum_{n=-\infty}^{+\infty} (g[n] - \delta[n])^2 = \sum_{n=-\infty}^{+\infty} \left\{ \frac{1}{A} \left(s[n] - \sum_{k=1}^P a_k s[n-k] \right) - \delta[n] \right\}^2$$

is minimized by taking partial derivatives with respect to each a_k and setting the respective equations to zero. (It is left as an exercise to show that these points correspond to maxima.) With respect to a given a_i , ε is therefore minimized by solving

$$\frac{\partial \varepsilon}{\partial a_i} = \sum_{n=-\infty}^{+\infty} 2 \left\{ \frac{1}{A} \left(s[n] - \sum_{k=1}^P a_k s[n-k] \right) - \delta[n] \right\} \left(-\frac{1}{A} s[n-i] \right) = 0, \quad i = 1, 2, \dots, p,$$

which expands to

$$\frac{\partial \varepsilon}{\partial a_i} = \sum_{n=-\infty}^{+\infty} s[n]s[n-i] - \sum_{n=-\infty}^{+\infty} \sum_{k=1}^P a_k s[n-k]s[n-i] - \sum_{n=-\infty}^{+\infty} A\delta[n]s[n-i] = 0, \quad i = 1, 2, \dots, p.$$

Noticing that $\sum_{n=-\infty}^{+\infty} A\delta[n]s[n-i] = As[-i]$, this simplifies to

$$\frac{\partial \varepsilon}{\partial a_i} = \sum_{n=-\infty}^{+\infty} s[n]s[n-i] - \sum_{k=1}^P a_k \sum_{n=-\infty}^{+\infty} s[n-k]s[n-i] - As[-i] = 0, \quad i = 1, 2, \dots, p.$$

Defining the deterministic autocorrelation of $s[n]$ as

$$\phi_s[m] \equiv \sum_{n=-\infty}^{+\infty} s[n+m]s[n],$$

we finally arrive at the set of linear equations:

$$\boxed{\sum_{k=1}^P a_k \phi_s[i-k] = \phi_s[i] - As[-i], \quad i = 1, 2, \dots, p}, \quad (3)$$

and for causal $s[n]$,

$$\boxed{\sum_{k=1}^P a_k \phi_s[i-k] = \phi_s[i], \quad i = 1, 2, \dots, p}. \quad (4)$$

Equations 3 and 4 are referred to as the *autocorrelation normal equations*, or the *Yule-Walker equations*. They are often formulated for causal $s[n]$ as the following matrix multiplication:

$$\begin{bmatrix} \phi_{ss}[0] & \phi_{ss}[1] & \cdots & \phi_{ss}[p-1] \\ \phi_{ss}[1] & \phi_{ss}[0] & \cdots & \phi_{ss}[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{ss}[p-1] & \phi_{ss}[p-2] & \cdots & \phi_{ss}[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \phi_{ss}[1] \\ \phi_{ss}[2] \\ \vdots \\ \phi_{ss}[p] \end{bmatrix}, \quad (5)$$

or equivalently,

$$T_p \alpha_p = r_p.$$

In practice, the remaining gain factor A is often chosen so that the total energy in $s[n]$ and $s'[n]$ are equal, *i.e.* so that $\sum_n s^2[n] = \sum_n s'^2[n]$.

Interestingly, these same equations are the solution to a closely-related statistical problem. If $s[n]$ is to be modeled as the output of an all-pole filter driven by zero-mean white noise $w[n]$, *i.e.*

$$w[n] \longrightarrow \boxed{\frac{A}{1 - \sum_{k=1}^p a_k z^{-k}}} \longrightarrow s'[n],$$

or

$$s'[n] = \sum_{k=1}^p a_k s'[n-k] + Aw[n],$$

the model parameters a_k prescribe an inverse whitening filter. Linear equations are therefore obtained by considering an inverse whitening scenario:

$$s[n] \longrightarrow \boxed{\frac{1}{A} [1 - \sum_{k=1}^p a_k z^{-k}]} \longrightarrow g[n],$$

and the a_k 's are chosen so that $g[n]$ approximates white noise. Specifically, our stochastic prediction error $e[n]$ is defined as

$$e[n] \equiv s[n] - \sum_{k=1}^p a_k s[n-k],$$

and we would like to minimize

$$\varepsilon = E [e^2[n]].$$

As it turns out, the solution of this stochastic formulation once again gives the equations described by Equation 4, where $\phi_s[m] = E [s[n+m]s[n]]$.

We'll discuss an efficient way to solve these equations in the next lecture.