EART266 Geological Signal Processing
Lecture #8
Power Spectral Density: 3, Maximum Entropy

The maximum entropy method, like Blackman-Tukey, is parametric. It assumes that the process we have sampled is AR(q). Unlike BT (and the periodogram), the maximum entropy (ME) method doesn't employ Fourier transforms. To understand how it is obtained requires a little bit of machinery we don't have yet. So that's where I will start.

Let's define a new variable z:

$$z = e^{-2\pi i f \Delta T}$$

By definition, z is complex. In form, it resembles the DFT kernel for frequency f and time step $n = 1$. We now generalize this idea by imagining that f (frequency) is also complex and that is spans the entire complex plane rather than just the Nyquist band of the real axis. In fact, the frequency band we are used to (real and bounded above and below by the Nyquist frequency) maps z onto the unit circle in the complex plane:

With this definition, the naïve periodogram estimate of the power spectrum is (to within some constant multiplier):

$$I(f) = \left| \sum_{n=0}^{P-1} h_n z^n \right|^2$$

This estimate (which is also part of the BT method) can have zeroes in the complex z plane (i.e., it can vanish for some values of complex f), but not poles (reciprocals of 0). People often refer to processes whose spectrum can be represented this way as <u>all-zeroes</u>. Moving average processes (MA(q)) have spectra composed of a finite sum of powers of z and hence are an example of an all-zeroes process.

AR(q) processes, on the other hand, are an example of an <u>all-poles</u> process. Their spectra is represented as:

$$I(f) = \frac{\alpha_0}{\left| 1 + \sum_{k=1}^{q} \alpha_k z^k \right|^2}$$

The "poles" are values of complex f that cause the denominator to equal zero. Since the process must have finite energy (if we can sample it!), these poles do not lie on the real frequency line.

Comparing these two forms, it is clear that the former should be better at represented smooth spectra whereas the latter will be better at representing spectra with "lines" (strong concentrations of power over a narrow frequency band). Both can model the other, but the all-zeroes form takes infinitely many terms to represent the spectrum of a all-poles process with a finite number of terms, and vice-versa.

For a MA(q) process, determining the power spectrum requires knowledge of the α's in the above equation. The method by which we estimate them is known as the Maximum Entropy Method (or MEM for short). Many of the ideas used by MEM are applicable elsewhere in this class, so I will work through some of them. But first, here's a quick and dirty explanation of MEM.

Rayleigh's principle (aka Wiener-Khinchin) states that the Fourier transform of the auto-covariance function is equal to the power spectrum. Let $C(k)$ be the auto-covariance at lag k. It's z transform (using z for the Fourier kernel) is:

$$\sum_{k=-q}^{q} C(k) z^k$$

We posit that:

$$\frac{\alpha_0}{\left|1+\sum_{k=1}^{q}\alpha_k z^k\right|^2} \approx \sum_{k=-q}^{q} C(k) z^k$$

over the range z^{-q} to z^q (coefficients of like powers must match). Here q is the assumed order of the AR process. Note that the rational function expansion (LHS) implies a correlation at all lags–lags the sampled time series doesn't constrain! The way it extrapolates it implies the least information, hence the name Maximum Entropy.

Okay, but matching powers of z, we can estimate the $q + 1$ α's and use them to compute the power spectrum (I) at any frequency in the Nyquist band. The question is: how the $%&# do we estimate the α's?

In short, we do this by solving:

$$\begin{pmatrix} C(0) & C(1) & \dots & C(q) \\ & & & \\ C(q) & C(q-1) & \dots & C(0) \end{pmatrix} \begin{pmatrix} 1 \\ \alpha_1 \\ \vdots \\ \alpha_q \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

We will see why later. The matrix is a covariance matrix. These are rife in signal processing. It has Toeplitz structure (constant values along diagonals). This facilitates solution and allows one to iteratively increase q easily.

To get the MEM power spectrum estimate we solve this system. To do so, we need to know α_0 whose value we set as:

$$\alpha_0 = \frac{1}{P}\sum_{n=0}^{P-1} h_n^2$$

This is the mean square of the sampled time series. Before deriving this MEM equations, let's take a look at some MEM power spectrum estimates.

Many examples follow on subsequent pages.

Okay, so where did the equations for the MEM spectrum come from?

The answer lies in linear filter theory. Imagine we have some signal h_n and we wish to find a second signal a_n such that:

$$h_n * a_n = \delta_n$$

i.e., that the convolution of the two signals takes on the value unity for a lag of 0 and is zero for all other lags. In this context, we say the *a* is a filter. In this case, the filter takes an extended signal (*h*) and compacts it into a single time step. In the frequency domain (using the convolution theorem) the above expression can be written:

$$H(f)A(f) = 1$$

given that the Fourier transform of a unit-amplitude spike at time 0 is unity. Written this way, it is obvious that:

$$A(f) = \frac{1}{H(f)}$$

provides the filter $A(f)$ given $H(f)$ (and provided $H(f) \neq 0$). Note that the power spectrum of *h* is (to within a constant):

$$I_H(f) = |H(f)|^2 = \frac{1}{|A(f)|^2}$$

Thus knowledge of *a* is equivalent to knowledge of the power spectrum of *h*.

Now let's assume that h_n is an AR(q) process. Theory tells us that the power spectrum of such a process is given by:

$$I_H(f) = \frac{1}{\left| \sum_{k=0}^{q} \alpha_k z^k \right|^2}$$

(Note that this is slightly different from what I wrote before, but functionally equivalent. The alphas are not identical however). This implies:

$$A(f) = \sum_{k=0}^{q} \alpha_k z^k$$

From this, the inverse Fourier transform is easily obtained:

$$a_n = \begin{cases} \alpha_n & 0 \le n \le q \\ 0 & else \end{cases}$$

Thus a is causal (no signal before time step 0) and finite in duration. For these reasons, it is dubbed a FIR (finite-impulse response) filter.

Using this, we now turn our attention back to the α's.

Assume we have $q + 1$ samples of h_n. Since a_n is 0 for $n < 0$, the convolution of h and a is:

Time Step	$h_n * a_n$	δ_n
$n = 0$	$h_0 a_0$	1
$n = 1$	$h_1 a_0 + h_0 a_1$	0
•	•	•
$n = q$	$h_q a_0 + h_{q-1} a_1 + \ldots + h_0 a_q$	0

This system of $q + 1$ linear equations can be rewritten as:

$$\begin{pmatrix} h_0 & 0 & \ldots & 0 \\ h_1 & h_0 & \ldots & 0 \\ \vdots & & & \\ h_q & h_{q-1} & \ldots & h_0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_q \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Solving this system provides the a's and hence the power spectrum of h.

This analysis assumed that h was AR(q) (with q known). What if q isn't known? What if h isn't AR? What if h has measurement error in it?

To partially answer these questions, we examine:

$$h_n * a_n = \delta_n + \varepsilon_n$$

where the ε's represent the errors in assuming that h is a AR(q) process. If we have sampled h for $P >> q$ time steps, then we can estimate the a's using:

$$\begin{pmatrix} h_0 & 0 & \dots & 0 \\ h_1 & h_0 & \dots & 0 \\ \vdots & & & \\ h_P & h_{P-1} & \dots & h_{P-q} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_q \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_P \end{pmatrix}$$

We don't know the ε's, so we can use Least Squares to estimate the a's that minimize the misfit (minimize the sum of squares of the ε's). In math, we minimize:

$$E^2 = \varepsilon^{\mathbf{T}} \varepsilon$$

where the vector of misfits $\boldsymbol{\varepsilon}$ is given by:

$$\varepsilon = \mathbf{Ha} - \delta$$

The least squares estimate is gotten by solving the least squares normal equations:

$$\mathbf{H}^T \mathbf{Ha} = \mathbf{H}^T \delta$$

Looking at $\mathbf{H}^T\mathbf{H}$, we find that as $P \to \infty$, it approximates the covariance matrix $P\mathbf{C}$. Thus the least-squares normal equations look like:

$$P \begin{pmatrix} C(0) & C(1) & \dots & C(q) \\ C(1) & C(0) & \dots & C(q-1) \\ \vdots & & & \\ C(q) & C(q-1) & \dots & C(0) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_q \end{pmatrix} = \begin{pmatrix} h_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

which is not vastly different from the equations I supplied for the MEM estimate. Ah, but close only counts in horseshoes and hand-grenades. To go the whole nine-yards involves more math than I care to work through. So I'll try to briefly explain what is needed and leave it at that.

What we have left out is actually quite important. We have ignored the fact that AR(q) h is continuously driven by white noise. To determine the coefficients of the AR(q) process (q unknowns) we need to know the white noise series. We don't. The official equations for MEM deal with this in very nifty fashion. By assuming that $P >> q$, we can approximate the auto-covariance of the white noise as a delta function. That approximation and some other fairly annoying mathematics gets us the official equations.