Lecture 14:

Reading: Bendat and Piersol, Ch. 5.1-5.2, with attention to cross-covariance and cross-spectrum

Recap

We've started looking at frequency-wavenumber spectra, and we studied a lot of images. The goal for today is to figure out how to do the actual calculations.

Basics.

Consider a data set $y(x,t)$ where $-x_f < x < x_f$ and $-T < t < T$ (where here we're using $\pm x_f$ for the end points in space, and following our previous examples $\pm T$ for the end points in time.) We know from our definition of the Fourier transform that we can represent y as

$$y(x,t) = \sum_{n=-\infty}^{\infty} a_n e^{i2\pi f_n t}, \tag{1}$$

or

$$y(x,t) = \sum_{m=-\infty}^{\infty} a_m e^{i2\pi k_m x}, \tag{2}$$

and by extension we can do this process in two dimensions:

$$y(x,t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_{nm} e^{i2\pi(f_n t + k_m x)}, \tag{3}$$

where

$$2\pi k_m = \frac{2\pi m}{2x_f} \tag{4}$$

$$2\pi f_n = \frac{2\pi n}{2T} \tag{5}$$

The corresponding spectral density estimate can be calculated from the squared coefficients:

$$\hat{E}(k_m, f_n) = \frac{|a_{nm}|^2}{\Delta k \Delta f} \tag{6}$$

Practicalities.

Suppose we take a time-space data set and Fourier transform it in both directions. Here are some issues that might concern us:

1. *Does it matter whether we Fourier transform time or distance first?* All other things being equal, no. Time and space are orthogonal, and one will not influence the other. To see this consider the following:

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y(x,t) e^{i2\pi kx}\, dx \right] e^{i2\pi ft}\, dt = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} y(x,t) e^{i2\pi ft}\, dt \right] e^{i2\pi kx}\, dx = \tag{7}$$

   (How you demean or detrend might matter, and you can ponder these issues.)

2. *Can we compute a frequency-wavenumber spectrum by Fourier tranforming a frequency spectrum in space?* No. If you look at the above equation, you'll see that you need the full complex structure of the Fourier transformed variables in order to determine the spectrum. You have to retain the phase information. If you computed spectra first, you would suppress some of the signal that you wanted.

3. *When we compute frequency spectra, we usually only plot positive frequencies? Does the same strategy apply for frequency-wavenumber spectra?* No, at a given frequency we can have forward and backward propagating signals, so we'll often want to plot a half plane for frequency with positive and negative wavenumbers, or vice versa.

4. *In practical terms, how do we implement this?* If you Fourier transform in time and space, you'll end up with a domain that should be structured as in Figure 1. But Matlab will of course line up the frequencies and wavenumbers starting with positive and then negative. You can use the Matlab command "fftshift" to swap the presentation around.

Covariance

Early in the quarter we discussed the variance, and we left for later the concept of correlation or covariance. If we want to compare two time series, we can compute the variance of one record relative to the other. Formally we can write:

$$\text{cov}(x, y) = \langle x(t)y(t)\rangle. \tag{8}$$

or in discrete terms

$$\text{cov}(x, y) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i. \tag{9}$$

For comparison purposes, we often normalize this to produce a correlation coefficient, which is normalized by the variance:

$$r = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 \frac{1}{N}\sum_{i=1}^{N} y_i^2}}. \tag{10}$$

(You might wonder how to judge whether a correlation coefficient is statistically significant. Correlation coefficients should have a Gaussian distribution, which means that cumulative distribution function will be an error function. We can use this to determine the correlation coefficient that we might expect from an equivalent number of random white noise variables:

$$\delta r = \text{erf}^{-1}(p)\sqrt{\frac{2}{N}} \tag{11}$$

where p is the significance level we want to consider, typically 0.95, and N is the effective number of degrees of freedom.)

Coherence

Coherence provides information that is analagous to a correlation coefficient for Fourier transforms. It tells us whether two series are statistically linked at any specific frequency. This can be important if we think that the records are noisy or otherwise uncorrelated at some frequencies, but that they also contain statistically correlated signals.

To compute coherence, first we need a cross-spectrum. (We looked at this in passing when we considered Parseval's theorem, but at that stage, I quickly set my different variables equal to each other.) Consider two time series $x(t)$ and $y(t)$:

$$x(t) = \sum_{n=-\infty}^{\infty} X_n e^{i2\pi f_n t} \tag{12}$$

$$y(t) = \sum_{n=-\infty}^{\infty} Y_n e^{i2\pi f_n t} \tag{13}$$

The the cross spectrum is computed in analogy with the spectrum:

$$\hat{S}_{XY}(f_m) = \frac{\langle X_m^* Y_m \rangle}{T} \tag{14}$$

The relationship between the cross-spectrum and the covariance is analogous to the relationship between the spectrum and the variance. There are some important details to notice.

1. The cross spectrum is complex, while the spectrum was real.
2. The cross spectrum is computed as an average of multiple spectral segments.
3. In our discrete Fourier transform, we should be normalizing by N, as always, but we're mostly concerned with relative values.

The cross-spectrum is complex, and when we use it we distinguish between the real and imaginary parts. The real part is called the "co-spectrum":

$$C(f_k) = \frac{1}{N} \Re \sum_{n=1}^{N} (X_k Y_k^*) \tag{15}$$

and the imaginary part is called the "quadrature spectrum"

$$Q(f_k) = \frac{1}{N} \Im \sum_{n=1}^{N} (X_k Y_k^*). \tag{16}$$

To determine the frequency-space relationship between two data sets x_n and y_n, we first divide them into segments and Fourier transform them, so that we have a set of X_k's and a set of Y_k's. When we computed spectra, we found the amplitude of each X_k and then summed over all our segments. Now we're going to do something slightly different. For each segment pair, we'll compute the product of X times the complex conjugate of Y: $X_k Y_k^*$. Then we'll sum over all the segments. In Matlab this becomes

```
sum(X.*conj(Y),2);
```

The corresponding amplitude is $\sqrt{C^2(f_k) + Q^2(f_k)}$. For comparison the spectra for X was:

$$S_{xx}(f_k) = \frac{1}{N} \sum_{n=1}^{N} X_k X_k^*, \tag{17}$$

and it was always real.

The coherence resembles a correlation coefficient. It's the amplitude squared divided by the power spectral amplitudes for each of the two components:

$$\gamma_{xy}^2(f_k) = \frac{C^2(f_k) + Q^2(f_k)}{S_{xx}(f_k)S_{yy}(f_k)}. \tag{18}$$

(Sometimes you'll see G_{xx}, G_{yy}, and G_{xy} in place of S_{xx}, S_{yy}, and S_{xy}. Bendat and Piersol define S to represent the two sided cross-spectra density and G to represent to represent one-sided spectra.) It's really important that your spectra are based on more than one segment, that is that N exceeds 1. If that weren't the case, you'd just have a single realization of each spectra, and the resulting squared coherence would be

$$\gamma_{xy}^2(f_k) = \frac{X(f_k)Y^*(f_k)X^*(f_k)Y(f_k)}{X(f_k)X^*(f_k)Y(f_k)Y^*(f_k)} = 1, \tag{19}$$

which is not a terribly informative result. When it's done properly, coherence measures how well different segments of x and y show the same type of relationship at a given frequency.

The phase $\phi(f_k) = \tan^{-1}(-Q(f_k)/C(f_k))$ tells us the timing difference between the two time series. If $\phi = 0$, changes in x and y happen at the same time. If $\phi = \pi$, then x is at a peak when y is at a trough. And a value of $\phi = \pi/2$ or $\phi = -\pi/2$ tells us that the records are a quarter cycle different.

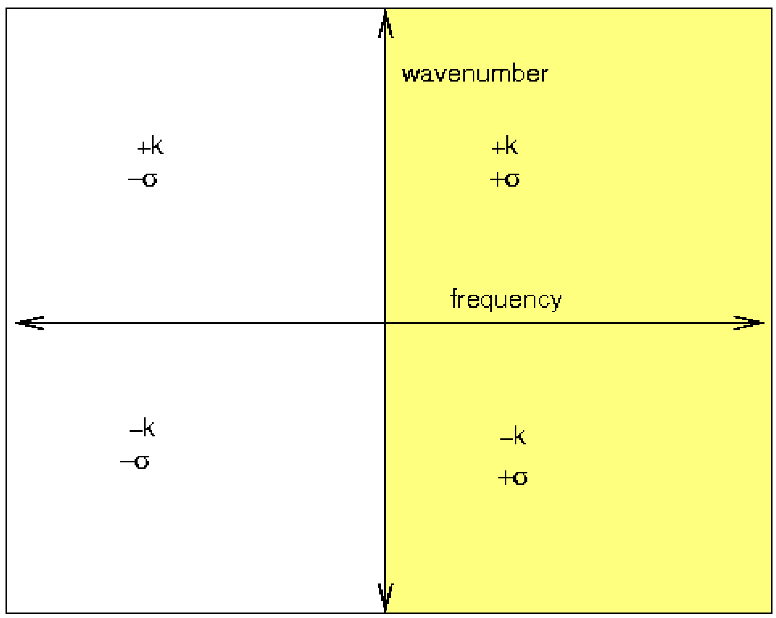

Figure 1: Fourier transform of time/space domain to form frequency/wavenumber domain. Note that $+k, +f$ is the complex conjugate of $-k, -f$ and similarly $+k, -f$ is the complex conjugate of $-k, +f$ so we normally plot only half the domain.