Lecture 6:

Reading: Bendat and Piersol, Ch. 2.1-2.2

Recap

Last time we looked at the least-squares fits and orthogonality of sines and cosines. This will lead us to the Fourier transform, which provides a way to re-represent data in terms of amplitudes of orthogonal sets of sines and cosines.

The Fourier Transform

So our least-squares fit of N data to N sinusoids was clearly too good to be true, but we're not doing fitting here, so we're going to proceed along this line of reasoning anyway. Our goal is to rerepresent all of the information in our data by projecting our data onto a different basis set. In this case we'll take the projection, warts and all, and we want to make sure we don't lose any information.

So we want to represent our data via sines and cosines:

$$x(t) = \frac{a_0}{2} + \sum_{q=1}^{\infty} \left(a_q \cos(2\pi q f_1 t) + b_q \sin(2\pi q f_1 t)\right), \tag{1}$$

where $f_q = 1/T_p$, and T_p is the duration of the record (following Bendat and Piersol). Formally we should assume that the data are periodic over the period T_p. We find the coefficients a and b by projecting our data onto the appropriate sines and cosines:

$$a_q = \frac{1}{T_p} \int_0^{T_p} x(t) \cos(2\pi q f_1 t)\, dt \tag{2}$$

and

$$b_q = \frac{1}{T_p} \int_0^{T_p} x(t) \sin(2\pi q f_1 t)\, dt \tag{3}$$

solved for $q = 0, 1, 2, ...$

It's not much fun to drag around these cosines and sines, so it's useful to recall that

$$\cos\theta = \frac{\exp(i\theta) + \exp(-i\theta)}{2} \tag{4}$$

$$\sin\theta = \frac{\exp(i\theta) - \exp(-i\theta)}{2i}, \tag{5}$$

which means that we could redo this in terms of $e^{i\theta}$ and $e^{-i\theta}$. In other words, we can represent our data as:

$$x(t) = \sum_{q=-\infty}^{\infty} [\hat{a}_q \exp(i2\pi q f_1 t)] = \sum_{q=-\infty}^{\infty} [\hat{a}_q \exp(i\sigma_q t)] \tag{6}$$

where $\sigma_q = 2\pi q/T$, and $\hat{a}_q$ represents a complex Fourier coefficient. If we solved for our coefficients for cosine and sine, then we can easily convert them to find the complex coefficients $\hat{a}_q$ for $\exp(i\sigma_q t)$ and $\exp(-i\sigma_q t)$. Consider :

$$a\cos\theta + b\sin\theta = \frac{a}{2}(e^{i\theta} + e^{-i\theta}) + \frac{b}{2i}(e^{i\theta} - e^{-i\theta}) \tag{7}$$

$$= \frac{a - ib}{2} e^{i\theta} + \frac{a + ib}{2} e^{-i\theta}. \tag{8}$$

This tells us some important things. The coefficients for $e^{i\theta}$ and $e^{i\theta}$ are complex conjugates. And there's a simple relationship between the sine and cosine coefficients and the $e^{\pm i\theta}$ coefficients. Instead of computing $\sum_{j=1}^{N} a_j \cos(\omega_j t)$ and $\sum_{j=1}^{N} b_j \sin(\omega_j t)$, we can instead find $\sum_{j=1}^{N} \hat{a}_j \exp(i\omega_j t)$ and then use the real and imaginary parts to represent the cosine and sine components. This gives us a quick shorthand for representing our results as sines and cosines.

---

sidebar: see minilecture
Fourier transform in continuous form

Bracewell's nice book on the Fourier transform refers to the data as $f(x)$ and its Fourier transform as $F(s)$, where x could be interpreted as time, for example, and s as frequency. Here I've rewritten to roughly use Bendat and Piersol's notation. In continuous form, the Fourier transform of $x(t)$ is $X(\omega)$ (where $\omega = qf_1$), and the process can be inverted to recover $x(t)$.

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi t\omega}\, dt \tag{9}$$

$$x(t) = \int_{-\infty}^{\infty} X(\omega) e^{i2\pi t\omega}\, d\omega \tag{10}$$

(following Bracewell).

But there are lots of alternate definitions in the literature:

$$X(\sigma) = \int_{-\infty}^{\infty} x(t) e^{-it\sigma}\, dt \tag{11}$$

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\sigma) e^{it\sigma}\, d\sigma \tag{12}$$

or

$$X(\sigma) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x(t) e^{-it\sigma}\, dt \tag{13}$$

$$x(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} X(\sigma) e^{it\sigma}\, d\sigma \tag{14}$$

So we always have to be careful about our syntax.

Given the vast array of notation, we're going to try very hard to stick to Bendat and Piersol's forms:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft}\, dt \tag{15}$$

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{i2\pi ft}\, df \tag{16}$$

The same questions about choices of notation apply in the discrete form that we consider when we analyze data.

Fourier transform in discrete form

We considered cosine and sine transforms, derived coefficients (a_q and b_q) for cosine and sine, and then showed that we could recombine these to make complex coefficients for $e^{i2\pi qf_1 t}$

and $e^{-i2\pi q f_1 t}$. We found these coefficients to be complex conjugates of each other. Since cosine/sine transformations and Fourier transforms using $e^{\pm i2\pi q f_1 t}$ are closely related, we can express results of one in terms of the other. In other words, instead of computing $\sum_{j=1}^{N} x_j \cos(2\pi f_j t)$ and $\sum_{j=1}^{N} x_j \sin(2\pi f_j t)$, we can instead find $\sum_{j=1}^{N} x_j \exp(i2\pi f_j t)$ and then use the real and imaginary parts to represent the cosine and sine components.

To reiterate, here's the Fourier transform in continuous form. Bendat and Piersol use the following:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t}\, dt \tag{17}$$

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{i2\pi f t}\, df \tag{18}$$

Data come in discrete form. If they are uniformly separated (in time or space), then they are easy to Fourier transform. The same questions about choices of notation apply in the discrete form that we consider when we analyze data. And we can get ourselves really confused. So we have to keep in mind one rule: we don't get to create energy. That means that we need to have the same total variance in our data set in the time domain as we have in the frequency domain. This is Parseval's theorem, and we'll return to it.

One of the glories of the Fourier transform is that we can take all of these projections and make them extremely efficient through the Fast Fourier Transform (FFT). In principle, FFT's are most efficient if you compute them for records that are a power of 2 in length, so 64 or 128 or 256 points for example. But modern FFTs are fast even if your data set doesn't have 2^n elements. Moreover, a year doesn't have 2^n days, so trying to force a data record to conform to a length of 2^n can suppress some of the natural periodicity.

Mathematically the Matlab definitions look like this:

$$X_k = \sum_{n=1}^{N} x_n \exp(-i2\pi(k-1)(n-1)/N), \tag{19}$$

where frequency labels k and data labels n go from 1 to N. Here capital letters are used to denote Fourier transformed variables. Matlab computes this using the command "fft".

The inverse of the Fourier transform is computed using "ifft" and is defined to be:

$$x_n = \frac{1}{N} \sum_{k=1}^{N} X_k \exp(i2\pi(k-1)(n-1)/N) \tag{20}$$

In Matlab the Fourier transform and inverse Fourier transform become:

```
f=fft(x)
x_new=ifft(f)
```

To make Parseval's theorem work, the variance of our data has to equal the variance of the Fourier transform. Thus we'll want to compare:

```
sum(x.^2)
sum(abs(f).^2)
f'*f
sum(f.*conj(f))
```

They don't quite agree, so we'll see that we should divide the Fourier transform by N, the number of data points.

What do we gain by Fourier transforming our data?

We live life in the time domain, so it's sometimes hard to think about the world as seen in the frequency domain. While linear trends aren't well represented by the Fourier transform, the Fourier tranform is particularly effective for representing sinusoidal oscillations. Solar radiation that warms the Earth varies on a 365.25 day cycle with the seasons, and on a 24 hour cycle, with the rising and setting of the sun. Ocean tides vary at semidiurnal (12.4 hour) and diurnal frequencies (as well as being modulated on fortnightly and monthly intervals.) Thus if you look at data from a tide gauge, you see oscillatory fluctuations at a variety of different frequencies, as shown in the slides. If we solve for the tidal amplitudes, we find for example:

Symbol	Frequency (cpd)	Amplitude (cm)	Greenwich Epoch
O1	0.92953571	8.91	217
P1	0.99726209	5.32	224
K1	1.00273791	16.12	225
M2	1.93227361	9.97	354
S2	2.00000000	6.45	357

The complex Fourier coefficients that emerge from the fft might seem confusing, but they give us a lot of information about our data, allowing us, for example to tell whether there is more energy at frequency σ_j compared with frequency σ_l. The Fourier coefficients are complex so this comparison might seem confusing, but we'll just examine the squared magnitudes of the coefficients: $|a_j|^2$.

Of course, if we knew the frequency exactly, we could just do a least-squares fit, but often we aren't exactly sure of the frequencies in question—there might be energy spread over a broad range of frequencies, and the Fourier tranform provides us with a way to examine our data in terms of oscillatory signals.

Questions about pier data

Here are some questions (yours plus some stray questions):

1. Why is the manual temperature warmer than the automated?
2. What is the depth of the automated sensor, and how does it compare with the manual sensor?
3. Could a mechanical system be set up to move with the water surface so that the manual and automated systems measured at the same depth?
4. Have there been procedural changes in the manual system over time?
5. Can manual tempartures be corrected to match the automated temperatures? Why are the two sensors different?
6. Why is the automated system set to sample at 4-minute intervals?
7. What happens when one of the sensors is down?
8. How do temperature and salinity vary seasonally, annually, interannually, and on decadal time scales? What accuracies do we need to assess these modes of variability?

9. How have changes in the location of the pier influenced the long-term record? What is the stability of the instrumentation? Do error bars evolve over time?
10. Does the pier create noise/flow distortion?
11. How well do the manual and automated records match?
12. Who quality controls the data, and how?
13. Is there visibility data?
14. Which of the variables associated with the pier record are measured directly, and which are inferred?
15. What methods have been used to collect measurements on the pier, and how consistent are they?
16. What variables are collected? What is the formal uncertainty? What is the sampling frequency?
17. When did automated sampling start? When did automated sampling start being reliable?
18. What are purposes of automated vs manual systems?
19. What level of adjustment is applied to make manual and automated data match?
20. What has been published about these data?
21. How often are the sensors serviced?
22. What accounts for gaps in the data?
23. How hands on is the automated system? What's really automatated?
24. What depths are the sensors?
25. Where are the sensors?
26. Are multiple sensors used are merged?
27. What is the local geographic variability?
28. What is the time of day of measurement? How clearly is that documented?
29. What does instrument failure look like in the data records?
30. What error flags are available for the data?
31. How long are records?
32. How is equipment calibrated? And how often?