

Lecture 5:

Reading: Bendat and Piersol, Ch. 4.5.1, 11.1, 11.2

Recap

Last time we looked at least-squares fitting. We derived the formula for a least-squares fit and showed that we could find a linear trend and a sinusoidal variation. We set up the least squares problem and looked at sinusoidal fits, and we finished up by asking what would happen if we fit multiple sinusoidal signals at once. That's where we'll start today.

Least-squares fits and misfit

You'll recall that last time we considered a least-squares fit of the form

$$\mathbf{Ax} + \mathbf{n} = \mathbf{y}. \quad (1)$$

The misfit is defined as a squared quantity so should follow the χ^2 statistic. (Yet another use of χ^2 .) If I believe my *a priori* uncertainties in my data are σ , then I expect that my misfit should roughly match my uncertainty so I can define a weighted summed misfit:

$$\chi^2 = \sum_{i=1}^N \left(\frac{y_i - \sum_{j=1}^M a_{i,j} x_j}{\sigma_i} \right)^2. \quad (2)$$

Here we're summing the squared misfit of each row in our matrix equation, weighted by our uncertainty. If our error bars make sense, then this should yield about N , reduced by the number of functions were fitting. So we expect that χ^2 will be about $N - M$, which is the number of degrees of freedom, ν . Formally we can decide if our fit is too good to be true by evaluating χ^2 using the incomplete gamma function, to find where the observed χ^2 falls within the expected pdf of a *chi*² distribution:

```
p=gammainc(chi_squared/2, nu/2, 'upper')
```

(Here we divide both χ^2 and ν by 2 because of the way that `gammainc` is defined.) If p is smaller than 0.05 or greater than 0.95, then our observed value is outside the range expected for 90% of observed χ^2 values. If p is near 1, it can tell us that our fit is too good to be true. Likewise, if p is too small, it can tell us that our fit isn't properly representing the data—either the model is wrong, or the *a priori* error bars are too large.

Alternatively, we can solve for upper and lower threshold χ^2 values: if $\chi^2 > \chi_{upper}^2$, then we reject the hypothesis at the α level. And if $\chi^2 < \chi_{lower}^2$ we also know we're out of range. So

```
chi2_upper = gammaincinv(0.05, nu/2, 'upper') * 2.
chi2_lower = gammaincinv(0.05, nu/2, 'lower') * 2.
```

What happens in the limit when we fit N data points with N columns in matrix \mathbf{A} ? The matrix \mathbf{A} is an $N \times N$ square matrix, and we are solving for as many unknowns in \mathbf{x} as we had data in \mathbf{y} . In this case, if χ^2 is zero, p will be 1, warning us that we're over-fitting our data. What happens to our noise \mathbf{n} ? By using N orthogonal functions, we obtain a perfect fit and the noise is zero. That's convenient, but it loses any information that we might have had about uncertainties in our data. If we made noisy measurements, we might not have any reason to expect a perfect fit, but we'll have one anyway.

Multiple oscillatory signals

You might wonder if you'll bias your results by fitting for all of the sinusoidal variability all at once. Usually, the answer is no. Assuming that your time series is long enough, sinusoidal frequencies are orthogonal to each other: there is no correlation between $\sin(2\pi t/T)$ and $\sin(2\pi 2t/T)$, just as there is no correlation between the sine and cosine components.

So we can take this to the maximum limit. Suppose we just want to fit sines and cosines to our data. How many frequencies can we fit? If I'm going to do this, then I'll need to make sure that each row of my matrix equation is linearly independent, which means that I'll want to make sure that each column of \mathbf{A} is orthogonal, so I can't choose frequencies that are too closely spaced.

Least-Squares Fitting Sines and Cosines

Least-squares fitting is particularly tidy when the functions that we use for our fit, the columns of our matrix \mathbf{A} , are completely orthogonal, because then the fit to one function has no impact on the fit to the other functions.

Consider the special case where the columns of \mathbf{A} are made up of sines and cosines, so

$$\mathbf{A} = \begin{bmatrix} 1 & \cos(\omega t) & \sin(\omega t) & \cos(2\omega t) & \sin(2\omega t) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \quad (3)$$

where $\omega = 2\pi/T$ and T is the total duration of the data record. The dot product of any two columns i and j of \mathbf{A} is zero if $i \neq j$. If I have data at N evenly spaced time increments, t_1, t_2, \dots, t_N , then this orthogonality property holds for all frequencies from ω through $N\omega/2$. Since I have a sine and cosine at each frequency (up to frequency $N\omega/2$ where sine might be zero at all points in time), this means that I can define a total of N independent orthogonal columns in \mathbf{A} .

On the other hand, if I define a column of \mathbf{A} to have a frequency $\omega/2$, it won't be orthogonal to my other functions over the range of this data. For example, between 0 and T , $\sin(\omega/2)$ varies from 0 to 1 to 0 and is always positive, meaning that it will be positively correlated with a constant. In fact, sines and cosines with frequencies that are ω multiplied by integers ranging from 0 to $N/2$ make a complete set that spans all space, and there are no additional N -element vectors that I can add to \mathbf{A} that would also be orthogonal to all other columns of \mathbf{A} .

The orthogonality of the columns of \mathbf{A} is really important. It means that my solution for x_1 is completely independent of my solution for x_2 . Here are some results for a set of 128 random numbers, b .

$$\hat{b} = -0.0629 - 0.0620 \cos(\omega t) - 0.1339 \sin(\omega t) \quad (4)$$

$$\hat{b} = -0.0629 - 0.0960 \cos(2\omega t) + 0.1117 \sin(2\omega t) \quad (5)$$

$$\hat{b} = -0.0629 - 0.0620 \cos(\omega t) - 0.1339 \sin(\omega t) - 0.0960 \cos(2\omega t) + 0.1117 \sin(2\omega t), \quad (6)$$

where \hat{b} is our fitted approximation to b . You can see that the amplitudes of $\cos(\omega t)$ and $\cos(2\omega t)$ are the same regardless of whether \mathbf{A} contains 3 columns or 5 columns.

Code to do this:

```
x=randn(128,1);
A=[ones(128,1) cos(2*pi*(1:128)'/128) sin(2*pi*(1:128)'/128) cos(4*pi*(1:128)'/128) sin(4*pi*(1:128)'/128)];
% solution for fitting 5 values
inv(A'*A)*A'*x
```

```
% solution for fitting 3 values, with low frequency sine and cosine
inv(A(:,1:3))' * A(:,1:3)) * A(:,1:3))' * x

% solution for fitting 3 values, with higher frequency sine and cosine
inv(A(:, [1 4:5]))' * A(:, [1 4:5])) * A(:, [1 4:5]))' * x
```

What happens if we want to fit the frequency $\omega/2$? In this case, it won't be orthogonal to my other functions over the range of this data. For example, between 0 and T , $\sin(\omega/2)$ varies from 0 to 1 to 0 and is always positive, meaning that it will be positively correlated with a constant. In fact, sines and cosines with frequencies that are ω multiplied by integers ranging from 0 to $N/2$ make a complete set that spans all space, and there are no additional N -element vectors that I can add to \mathbf{A} that would also be orthogonal to all other columns of \mathbf{A} .

If we take a time series of N elements, then the lowest frequency that we can resolve is 1 cycle per N elements, so $\cos(2\pi i/N)$, where our counter i runs from 1 to N (or from 0 to $N - 1$). We can find two coefficients for this: one for the \cos component and one for the \sin component.

Actually, maybe a better way to think about this is that the lowest frequency we can resolve is $\cos(0i/N) = 1$, which is a constant and represents the mean. Since $\sin(0) = 0$, there is only a cosine component for frequency 0.

At any rate, after considering 1 cycle per N points, the next frequency we can resolve that will actually be fully orthogonal is 2 cycles per N points. We can keep counting upward: 3 cycles per N points, 4 cycles per N points, and so forth. All of these are guaranteed to be orthogonal over our domain of N points.

What is the maximum number of cycles that we can resolve in N points? One possibility would be that the maximum is N cycles per N points. That would require a full sinusoidal oscillation squeezed between data element 1 and data element 2. But if you think about it, we wouldn't expect to have enough information to determine the amplitude of a sine wave that had to squeeze itself between consecutive observations. Moreover if N cycles per N points were the maximum, this would mean that we'd be solving for $2N$ coefficients with only N data points. Clearly that would require more information than we have available.

Last time we noted that with N data points, we can fit a maximum of N functions. If we fit sine and cosine pairs, this translates into $N/2$ cosines and $N/2$ sines. The highest frequency we can resolve is 1 cycle every 2 data points, so $N/2$ cycles in N points, and this is the *Nyquist frequency*.

And the strategy of least-squares fitting all possible frequencies that can be resolved represents the *discrete Fourier transform*. It's a slow and inefficient Fourier transform, but it is the essence of this class and it will be the building block for everything we do in the remainder of the quarter.

Orthogonality and Sines and Cosines

Last time we talked about the importance of having independent columns in our matrix \mathbf{A} and noted that sines and cosines are particularly useful since they are orthogonal. Let's work through this a little more carefully.

Consider a record of duration T with N data points. I can imagine squeezing into the period

T , one sine wave, or two, or three, or four. How do I tell if my records are orthogonal?

$$\begin{aligned} \int_0^T \sin\left(\frac{2\pi nt}{T}\right) \sin\left(\frac{2\pi mt}{T}\right) dt &= \frac{1}{2} \int_0^T \cos\left(\frac{2\pi(n-m)t}{T}\right) - \cos\left(\frac{2\pi(n+m)t}{T}\right) dt \\ &= \frac{1}{2} \frac{T}{2\pi} \left[\frac{\sin\left(\frac{2\pi(n-m)t}{T}\right)}{n-m} - \frac{\sin\left(\frac{2\pi(n+m)t}{T}\right)}{(n+m)} \right] \Bigg|_0^T \\ &= \begin{cases} 0, & \text{if } n \neq m \\ \frac{T}{2}, & \text{if } n = m \end{cases} \end{aligned}$$

(What matters is that this is only non-zero in the special case when $n = m$. For the moment, the fact that the integral yields $T/2$ when $n = m$ is a minor detail.) By extension the same applies for two cosines, or a sine multiplied by a cosine.

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:

$$x(t) = \frac{a_0}{2} + \sum_{q=1}^{\infty} (a_q \cos(2\pi q f_1 t) + b_q \sin(2\pi q f_1 t)), \quad (7)$$

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:

$$a_q = \frac{1}{T_p} \int_0^{T_p} x(t) \cos(2\pi q f_1 t) dt \quad (8)$$

and

$$b_q = \frac{1}{T_p} \int_0^{T_p} x(t) \sin(2\pi q f_1 t) dt \quad (9)$$

solved for $q = 0, 1, 2, \dots$

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

$$\cos \theta = \frac{\exp(i\theta) + \exp(-i\theta)}{2} \quad (10)$$

$$\sin \theta = \frac{\exp(i\theta) - \exp(-i\theta)}{2i}, \quad (11)$$

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:

$$x(t) = \sum_{q=-\infty}^{\infty} [\hat{a}_q \exp(i2\pi q f_1 t)] = \sum_{q=-\infty}^{\infty} [\hat{a}_q \exp(i\sigma_q t)] \quad (12)$$

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 :

$$a \cos \theta + b \sin \theta = \frac{a}{2}(e^{i\theta} + e^{-i\theta}) + \frac{b}{2i}(e^{i\theta} - e^{-i\theta}) \quad (13)$$

$$= \frac{a - ib}{2}e^{i\theta} + \frac{a + ib}{2}e^{-i\theta}. \quad (14)$$

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.

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 $x(t)$.

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi t\omega} dt \quad (15)$$

$$x(t) = \int_{-\infty}^{\infty} X(\omega)e^{i2\pi t\omega} d\omega \quad (16)$$

(following Bracewell).

But there are lots of alternate definitions in the literature:

$$X(\sigma) = \int_{-\infty}^{\infty} x(t)e^{-it\sigma} dt \quad (17)$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\sigma)e^{it\sigma} d\sigma \quad (18)$$

or

$$X(\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t)e^{-it\sigma} dt \quad (19)$$

$$x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X(\sigma)e^{it\sigma} d\sigma \quad (20)$$

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:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft} dt \quad (21)$$

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{i2\pi ft} df \quad (22)$$

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.