Lecture 12:

Recap

We've looked at several strategies for computing spectra and estimating degrees of freedom—segmenting data and computing the periodogram (the Fourier transform) or computing the periodogram (Fourier transform) and then averaging. Now, let's look at this from a different angle by considering the covariance.

Using the auto-covariance to think about spectra.

Now let's look at spectra from a different perspective. When we talked about Parseval's theorem, we took a look at autocovariance. That was the convolution of $x(t)$ with its time reversal, $x(-t)$.

$$y(\tau) = \int_{-\infty}^{\infty} x(t)x(\tau + t)dt. \tag{1}$$

More formally, we might write this autocovariance as $R_{xx}(\tau)$.

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(\tau + t)dt. \tag{2}$$

Now, what if we Fourier transform R?

$$S_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau)e^{-i2\pi f\tau}\, d\tau. \tag{3}$$

Formally, this and its inverse transform are the Wiener-Khinchine relations.

Now let's think about starting with two functions, $x(t)$ and $y(t)$. We can write their Fourier transforms:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}\, dt \tag{4}$$

$$Y(f) = \int_{-\infty}^{\infty} y(t)e^{-i2\pi ft}\, dt. \tag{5}$$

So now let's define X times the complex conjugate of Y. (Why do we consider the complex conjugate? Because it's how we always multiply vectors.) We find the Fourier transform of the complex conjugate by substituting $-i$ for i everywhere it appears:

$$Y^*(f) = \int_{t=-\infty}^{\infty} y(t)e^{i2\pi ft}\, dt \tag{6}$$

$$= \int_{-t=-\infty}^{\infty} y(-t)e^{-i2\pi ft}\, d(-t) \tag{7}$$

$$= \int_{t=\infty}^{-\infty} -y(-t)e^{-i2\pi ft}\, dt \tag{8}$$

$$= \int_{t=-\infty}^{\infty} y(-t)e^{-i2\pi ft}\, dt. \tag{9}$$

So

$$X(f)Y^*(f) = \int_{-\infty}^{\infty} k(t)e^{-i2\pi ft}\, dt. \tag{10}$$

For the moment, we have no idea what $k(t)$ should be, but we should be able to figure it out. If $X(f)Y^*(f)$ is a product in the frequency domain, then $k(t)$ should be a convolution in the time domain:

$$k(t) = \int_{-\infty}^{\infty} x(u)y(u-t)\,du. \tag{11}$$

(Remember that we'd normally use $t-u$ for a convolution; but here we reverse the sign to be consistent with using the complex conjugate $Y^*(f)$.) We used a derivation very similar to this in about Lecture 7, when we wanted to persuade ourselves that Parseval's theorem would work. But now we revisit this with a goal of looking closely at this convolved quantity, which represents the autocovariance. We can plug $k(t)$ into our equation to check this.

$$\begin{aligned}
\int_{-\infty}^{\infty} k(t)e^{-i2\pi ft}\,dt &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} x(u)y(u-t)\,du\right\}e^{-i2\pi ft}\,dt & (12)\\
&= \int_{-\infty}^{\infty}\int_{u-t=\infty}^{-\infty} -x(u)y(u-t)e^{i2\pi f(u-t)}e^{-i2\pi fu}\,d(u-t)\,du & (13)\\
&= \int_{-\infty}^{\infty} x(u)e^{-i2\pi fu}\left\{\int_{-\infty}^{\infty} y(u-t)e^{i2\pi f(u-t)}\,d(u-t)\right\}du & (14)\\
&= \int_{-\infty}^{\infty} x(u)e^{-i2\pi fu}Y^*(f)\,du & (15)\\
&= Y^*(f)\int_{-\infty}^{\infty} x(u)e^{-i2\pi fu}\,du & (16)\\
&= X(f)Y^*(f). & (17)
\end{aligned}$$

Here we've taken advantage of the fact that the integral runs from $-\infty$ to $+\infty$ which lets us treat $u-t$ as a variable that depends only on t.

So we can think about what happens when $x(t) = y(t)$, so that

$$k(t) = \int_{-\infty}^{\infty} x(u)x(u-t)\,du = R_{xx}(-t). \tag{18}$$

This means that $k(t)$ is the autocovariance of x. The autocovariance is symmetric, so we could also write this as

$$k(t) = \int_{-\infty}^{\infty} x(u+t)x(u)\,du = R_{xx}(t). \tag{19}$$

Regardless

$$|X(f)|^2 = \int_{-\infty}^{\infty} k(t)e^{-i2\pi ft}\,dt. \tag{20}$$

This says that the Fourier transform coefficients squared (what we use when we compute spectra) are equivalent to the Fourier transform of the autocovariance.

Using the auto-covariance to compute spectra requires averaging, just as we did by segmenting our data and using the fft, but there's one tidy little trick. Let's use some white noise again, and take a look at our options:

1. Suppose we start with a big matrix of white noise, and we compute the autocovariance for each column of our matrix, then Fourier transform, and use these to compute a spectrum. We'll end up doing something along these lines:

```
A=randn(1000,100);
for i=1:100
 AcA(:,i)=xcov(A(:,i),A(:,i),'unbiased');  % autocovariance for
end
fAcA=fft(AcA(500:1500,:)); % Fourier transform of autocovariance
frequency=(0:500)/1000;
loglog(frequency,abs(mean(fAcA(1:501,:),2)),'LineWidth',3)
set(gca,'FontSize',16)
xlabel('Frequency (cycles per data point)','FontSize',16)
ylabel('Spectral energy','FontSize',16)
```

2. Alternatively, we could average all of the autocovariances, and then Fourier transform:

```
mean_AcA=mean(AcA,2);
fmean_AcA=fft(mean_AcA(500:1500));
hold on
loglog(frequency,abs(fmean_AcA(1:501,:))*1.1,'r','LineWidth',3)
legend('average of FFTs of many autocovariances',...
     'FFT of averaged autocovariance (scaled by 1.1)')
```

3. For comparison, the peridogram-based determined from the fft of the data:

```
fA=fft(A);
ampA=abs(fA(1:501,:)).^2/1000; ampA(2:500,:)=2*ampA(2:500,:);
loglog(frequency,mean(ampA,2),'LineWidth',3)
```

In the results, shown in Figure 1, the curves are identical, though the red line has been scaled up by 10% to make both visible. There are some normalizations here that we haven't properly confronted. Notably we are missing factor of 2 for the autocovariance cases. As with the simple Fourier transform (which we can properly refer to as a "periodogram"), we're only plotting half the spectrum, so we need double the energy. If we were dealing with real data, we'd also need a factor of N or Δt to properly normalize our fft. Details can be sorted out later, and Thomson and Emery provide a bit of guidance on this.

You shouldn't be surprised that averaging before or after the FFT leads to the same results, since averaging has no impact on the FFT. But this might give you an idea of how you can take advantage of the autocovariance to compute spectra from gappy data.

All of this means that we could compute spectra without needing to chunk our data and compute lots of ffts, provided that we had a good estimate of the autocovariance. In the days before the development of the FFT, the autocovariance was a natural pathway to determining the spectrum, since it was clean and easy to compute. And now, with modern computing, you might not feel like there's any need to take advantage of the FFT anymore. If you can obtain the best possible estimate of the autocovariance, by whatever means necessary, then you should be able to compute one FFT and obtain reasonable estimate of the spectrum, without concern for data gaps or computational speed.

Formally if you compute spectra from the autocovariance, you need to think about averaging just as thoroughly as you do for segmented windowed data (through the Welch method) or for frequency-averaged spectra (the Daniell method). In this case, the challenge comes in deciding

what fraction of the autocovariance to actually Fourier transform. If you use the autocovariance over the entire data range, your autocovariance estimator has too much uncertainty at the large lags, and the resulting spectrum will have large uncertainties. Textbook look up tables advise on degrees of freedom associated with each of these options.

Autocovariance in discrete form.

Last time we went through the representation of the autocovariance using an integral form. Let's rewrite this in terms of the discrete Fourier transform. In this case, the mean of our data is:

$$\langle x \rangle = \frac{1}{2T} \int_{-T}^{T} x(t) e^{i0} \, dt = a_0. \tag{21}$$

and the variance is

$$\langle x * x \rangle = \frac{1}{2T} \int_{-T}^{T} x^*(t) x(t) \, dt - |a_0|^2. \tag{22}$$

We use the complex conjugate here, just in case $x(t)$ is represented as a complex number, since this will give us the sum of the squares. Notice that we've remembered to subtract out the mean (our frequency zero Fourier coefficient).

In similar notation, we can write the covariance (for finite record length $2T$) as:

$$R(\tau) = \frac{1}{2T} \int_{-T}^{T} x^*(t) x(t+\tau) \, dt - |a_0|^2. \tag{23}$$

This lets us write out an expression for the variance R in terms of the discrete Fourier coefficients:

$$\begin{aligned} R(\tau) &= \frac{1}{2T} \int_{-T}^{T} \left[\sum_{n=-\infty}^{\infty} a_n^* e^{-i2\pi f_n t} \sum_{m=-\infty}^{\infty} a_m e^{i2\pi f_m (t+\tau)} \right] dt - |a_0|^2 & (24) \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_n^* a_m e^{+i2\pi f_m \tau} \frac{1}{2T} \int_{-T}^{T} e^{i(-2\pi f_n + 2\pi f_m)t} \, dt - |a_0|^2 & (25) \\ &= \sum_{m=-\infty}^{\infty} |a_m|^2 e^{+i2\pi f_m \tau} - |a_0|^2 & (26) \end{aligned}$$

where we used a Kronecker delta (δ_{nm}) to eliminate the integral with $e^{2\pi i(-f_n + f_m)t}$ except when $n = m$, and we subtracted a_0^2 at the end to match our original definition. In setting this up, recall (from lecture 5) that the Fourier transform uses $e^{-i2\pi ft}$, and the inverse transform uses $e^{+i2\pi ft}$. We're using the inverse transform here (though the signs reversed when we had the complex conjugate. This tells us that the Fourier transform of the autocovariance can be expressed by the squared Fourier coefficients. (So we could avoid the Fourier transform completely and just work with the auto-covariance.)

In this form, Parseval's theorem simply says that

$$\begin{aligned} R(0) &= \frac{1}{2T} \int_{-T}^{T} x^*(t) x(t+0) \, dt - |a_0|^2 & (27) \\ &= \sum_{m=-\infty}^{\infty} |a_m|^2 - |a_0|^2 & (28) \end{aligned}$$

meaning that the variance of x is the sum of magnitudes of the Fourier coefficients.

In the case that we considered last time, with white noise, let's consider the autocovariance:

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(\tau + t)dt = \begin{cases} 0 & \text{for } \tau \neq 0 \\ 1 & \text{for } \tau = 0 \end{cases} \tag{29}$$

This is true, because white noise is uncorrelated except at zero lag.

Alternatively, if I consider red noise, then the noise will be correlated from point to point, and the autocovariance will have a bit of spread. We can test this out:

```
% define red data with autoregressive process
a=randn(10000,1);
b(1)=a(1);
for i=2:length(a)
  b(i)=b(i-1)+a(i);
end

BB=xcorr(b,b)/max(xcorr(b,b));
BB_unbiased=xcorr(b,b,'unbiased')/max(xcorr(b,b,'unbiased'));
plot(-999:999,[BB' BB_unbiased'],'LineWidth',3)
```

Sidebar: Getting the units right

We've talked about units for spectra, but let's lay everything out in one place. Our fundamental principle is that we want Parseval's theorem to work. But this gets a tiny bit messy when we average multiple frequencies. Still the basic rule of Parseval's theorem is not that the sum of the squares equals the sum of the squared Fourier coefficients, but rather that the integrated variance equals the integral under the spectrum.

When Matthew lectured, his notes had the spectral estimate divided by N^2 and then divided by $df = 1/T$, so he was multiplying by T/N^2. But my sample code so far has only had a division by N. Why?

In the discrete Fourier transform, we have:

$$\sum_{n=0}^{N-1} x_n^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2, \tag{30}$$

or after spectral normalizations, maybe we write

$$\sum_{n=0}^{N-1} x_n^2 = \frac{1}{N} \left[|X_0|^2 \, \Delta f + \sum_{k=1}^{(N/2-1} 2|X_k|^2 \, \Delta f + |X_{N/2}|^2 \, df \right] \tag{31}$$

$$\approx \frac{1}{N} \sum_{k=0}^{(N-1)/2} 2|X_k|^2. \tag{32}$$

For this discussion, to keep the equations compact, we'll use the approximation in the final line, neglecting the fact that the mean and the $k = N/2$ value should $k = N/2$ value should not be doubled. Many times we work with this form, since it gives us meaningful spectral slopes. But in continuous form, we had a form more like this:

$$\int_{-\infty}^{\infty} x^2(t)dt = \int_{-\infty}^{\infty} |X(f)|^2 \, df \tag{33}$$

where f is frequency in cycles per unit time (or we might use $\sigma = 2\pi f$ sometimes, where σ is frequency in radians per unit time.) If we want this integral form to work for our real data, then we have to be a bit careful with our normalizations. We're going to want the area under the curve in our spectrum to be equal to the total variance integrated over time. So if total integrated variance is

$$variance = \sum_{n=0}^{N-1} x_n^2 \, \Delta t \tag{34}$$

where $\Delta t = T/N$. Then the integrated spectrum should be

$$variance = \frac{\alpha}{N} \sum_{k=0}^{N/2-1} 2|X_k|^2 \, \Delta f, \tag{35}$$

where $\Delta f = 1/(N\Delta t) = 1/T$, and we'll need to figure out α to ensure that the spectral estimator that we compute still properly adheres to Parseval's theorem. This implies that we might imagine normalizing our spectra to have:

$$\sum_{n=0}^{N-1} x_n^2 \Delta t = \frac{\alpha}{N} \sum_{k=0}^{N/2-1} 2|X_k|^2 \Delta f = \frac{\Delta t}{N\Delta f} \sum_{k=0}^{N/2-1} 2|X_k|^2 \Delta f = \frac{T^2}{N^2} \sum_{k=0}^{N/2-1} 2|X_k|^2 \Delta f \tag{36}$$

Or maybe this makes for units that aren't easily compared, so we could normalize our spectra to represent the average energy per unit time in the time domain, and adjust the frequency domain accordingly:

$$\begin{aligned}
\frac{1}{T} \sum_{n=0}^{N-1} x_n^2 \Delta t &= \frac{1}{N\Delta t} \sum_{n=0}^{N-1} x_n^2 \Delta t && (37)\\
&= \frac{T}{N^2} \sum_{k=0}^{N/2-1} 2|X_k|^2 \Delta f && (38)\\
&= \frac{(\Delta t)}{N} \sum_{k=0}^{N/2-1} 2|X_k|^2 \Delta f && (39)\\
&= \frac{1}{N^2 \Delta f} \sum_{k=0}^{N/2-1} 2|X_k|^2 \Delta f && (40)\\
&= \frac{1}{N 2 f_{Nyquist}} \sum_{k=0}^{N/2-1} 2|X_k|^2 \Delta f && (41)
\end{aligned}$$

so we could divide our spectrum by twice the Nyquist frequency to have energy in units appropriate for comparing if we wanted to have our integrals match.

This isn't always the way we think about this, but it serves as our reminder that we should think about the units of our spectrum. What we know is that integral of our spectrum over a certain frequency range should give a measure of the signal variance:

$$\text{variance in a band} = \int_{f-\Delta f/2}^{f+\Delta f/2} |X(f)|^2 \, df \tag{42}$$

So if we expand this out, this implies that the units of $|X(f)|^2$ should be equivalent to variance divided by frequency, so it's our reminder that we'll label the y-axis units as the squared units of x divided by frequency, with a normalization to account for the units of time in our data.

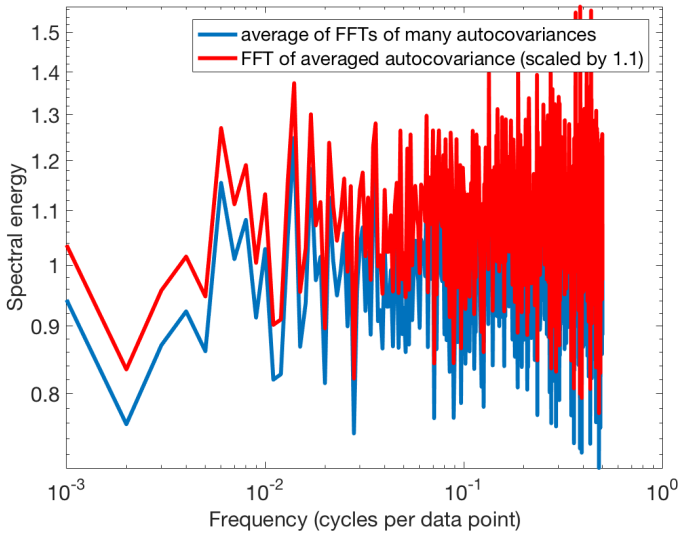

Figure 1: Spectra for white noise, computed by Fourier transforming 100 realizations of the autocovariance function (blue), or by Fourier transforming a smoothed autocovariance function computed from 100 realizations of the data (red). The red line is scaled upward by a factor of 1.1.