2.161 Signal Processing: Continuous and Discrete
Fall 2008

Massachusetts Institute of Technology
Department of Mechanical Engineering

2.161 *Signal Processing - Continuous and Discrete*
Fall Term 2008

Lecture 23[1]

Reading:

- Proakis and Manolakis: Secs. 14.1 – 14.2
- Oppenheim, Schafer, and Buck: 10.6 – 10.8
- Stearns and Hush: 15.4, 15.6

1 Non-Parametric Power Spectral Density Estimation

In Lecture 22 we defined the *power-density spectrum* $\Phi_{ff}(\mathrm{j}\,\Omega)$ of an infinite duration, real function $f(t)$ as the Fourier transform of its autocorrelation function $\phi_{ff}(\tau)$

$$\Phi_{ff}(\mathrm{j}\,\Omega) = \mathcal{F}\left\{ \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} f(t)f(t+\tau)\,\mathrm{d}t \right\}.$$

with units of (physical-units)^2.s or (physical-units)^2/Hz, where physical-units are the units of $f(t)$, for example volts. The waveform power contained in the spectral region between $\Omega_1 < |\Omega| < \Omega_2$ is

$$\begin{aligned} \mathbf{P} &= \frac{1}{2\pi}\left(\int_{-\Omega_2}^{-\Omega_1} \Phi_{ff}(\mathrm{j}\,\Omega)\,\mathrm{d}\Omega + \int_{\Omega_1}^{\Omega_2} \Phi_{ff}(\mathrm{j}\,\Omega)\,\mathrm{d}\Omega \right) \\ &= \frac{1}{\pi}\int_{\Omega_1}^{\Omega_2} \Phi_{ff}(\mathrm{j}\,\Omega)\,\mathrm{d}\Omega \end{aligned}$$

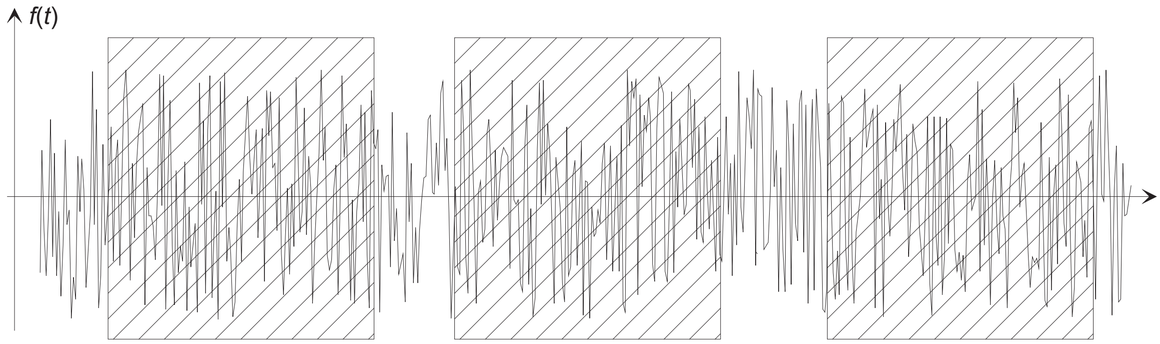
since $\Phi_{ff}(\mathrm{j}\,\Omega)$ is a real, even function.

Similarly, we defined the *energy-density* spectrum $R_{ff}(\mathrm{j}\,\Omega)$ of a real finite duration waveform $f(t)$ of duration T as the Fourier transform of its energy based autocorrelation function $\rho_{ff}(\tau)$

$$R_{ff}(\mathrm{j}\,\Omega) = \mathcal{F}\left\{ \int_0^T f(t)f(t+\tau)\,\mathrm{d}t \right\}.$$

with units of (physical-units)^2.s^2 or (physical-units)^2.s/Hz, again where physical-units are the units of $f(t)$.

In this lecture we address the issue of *estimating* the PSD (power spectral density) $\Phi_{ff}(\mathrm{j}\,\Omega)$ of an infinite process using a finite length sample of the process. PSD analysis is an important tool in engineering analysis. The practical problem is to form *reliable* estimates of the PSD from *finite* records of an infinite random process. For example, the following figure shows a stationary random process with three possible finite records. An estimator $\hat{\Phi}_{ff}(\mathrm{j}\,\Omega)$ of $\Phi_{ff}(\mathrm{j}\,\Omega)$ is to made from one of the finite-length records.



We ask ourselves about the statistics of estimators derived from the different records, in particular,

(1) the *bias* in the estimator

$$B\left(\hat{\Phi}_{ff}(\mathrm{j}\,\Omega)\right) = \Phi_{ff}(\mathrm{j}\,\Omega) - \mathcal{E}\left\{\hat{\Phi}_{ff}(\mathrm{j}\,\Omega)\right\}$$

(2) the *variance* of the estimator

$$V\left(\hat{\Phi}_{ff}(\mathrm{j}\,\Omega)\right) = \mathcal{E}\left\{\left(\Phi_{ff}(\mathrm{j}\,\Omega) - \hat{\Phi}_{ff}(\mathrm{j}\,\Omega)\right)^2\right\}$$

1.1 The Periodogram

1.1.1 The Continuous Periodogram

If $f(t)$ is a stationary, real, random process, its autocorrelation function is defined by the ensemble statistics

$$\phi_{ff}(\tau) = \mathcal{E}\left\{f(t)f(t+\tau)\right\}.$$

For an ergodic process, if we have a single record of duration T we can compute an estimator, $\hat{\phi}_{ff}(\tau)$ based on the time average:

$$\hat{\phi}_{ff}(\tau) = \frac{1}{T}\int_{-T/2}^{T/2} f(t)f(t+\tau)\,\mathrm{d}t,$$

and

$$\phi_{ff}(\tau) = \lim_{T\to\infty} \hat{\phi}_{ff}(\tau).$$

Furthermore, the Fourier transform of $\hat{\phi}_{ff}(\tau)$ provides an estimator $\hat{\Phi}_{ff}(\mathrm{j}\,\Omega)$ of the PSD

$$\begin{aligned}
\hat{\Phi}_{ff}(\mathrm{j}\,\Omega) &= \int_{-T/2}^{T/2} \hat{\phi}_{ff}(\tau)\, \mathrm{e}^{-\mathrm{j}\,\Omega\tau}\, \mathrm{d}\tau \\
&= \frac{1}{T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} f(t)f(t+\tau)\,\mathrm{d}t\, \mathrm{e}^{-\mathrm{j}\,\Omega\tau}\, \mathrm{d}\tau \\
&= \frac{1}{T}F(\mathrm{j}\,\Omega)F(-\mathrm{j}\,\Omega) \\
&= \frac{1}{T}\left|F(\mathrm{j}\,\Omega)\right|^2 .
\end{aligned}$$

where $F(\mathrm{j}\,\Omega)$ is the Fourier transform of the finite data record. The *periodogram* estimator $I_T(\mathrm{j}\,\Omega)$ is then defined as

$$\boxed{I_T(\mathrm{j}\,\Omega) = \hat{\Phi}_{ff}(\mathrm{j}\,\Omega) = \frac{1}{T}\left|F(\mathrm{j}\,\Omega)\right|^2 .}$$

1.1.2 The Discrete-Time Periodogram

For an infinite sampled data record $\{f_n\}$, the autocorrelation function is defined by the expectation

$$\phi_{ff}(m) = \mathcal{E}\left\{f_n f_{n+m}\right\},$$

and, invoking ergodicity, a time-average definition is

$$\phi_{ff}(m) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f_n f_{n+m}.$$

As in the continuous case, we can use a single finite length record, of length N, to form an estimator of $\Phi_{ff}(\mathrm{j}\,\Omega)$. The equivalent periodogram definition (through the DFT) is

$$\boxed{I_N(k) = \hat{\Phi}_{ff}(k) = \frac{1}{N}F_kF_{-k} = \frac{1}{N}\left|F_k\right|^2 .}$$

where $\{F_k\}$ is the DFT of $\{f_n\}$.

1.1.3 The Relationship between $I_T(\mathrm{j}\,\Omega)$ and $I_N(k)$

We frequently want to used discrete-time analysis to estimate the PSD of a continuous waveform. Consider a finite sample set $\{f_n\}$, of length N, derived from sampling a continuous waveform $f(t)$ with sampling interval Δ, so that

$$f_n = f(n\Delta), \qquad n = 0,\ldots,N-1.$$

From sampling theory $F^*(\mathrm{j}\,\Omega)$, the Fourier transform of the sampled waveform $f^*(t)$ is

$$F^*(\mathrm{j}\,\Omega) = \Delta F(\mathrm{j}\,\Omega),$$

and through the DTFT

$$F^*(\mathrm{j}\,\Omega) = \sum_{n=0}^{N-1} f_n\, \mathrm{e}^{-\mathrm{j}\,n\Omega\Delta}.$$

Then the continuous periodogram, with record length $T = N\Delta$, evaluated at $\Omega = 2\pi k/N$ is

$$\begin{aligned} I_T\left(\frac{\mathrm{j}\,2\pi k}{T}\right) &= \frac{1}{T}\left|F\left(\frac{\mathrm{j}\,2\pi k}{N\Delta}\right)\right|^2 \\ &= \frac{1}{N\Delta}\left|\Delta F^*\left(\frac{\mathrm{j}\,2\pi k}{N}\right)\right|^2 \end{aligned}$$

and since

$$F^*\left(\frac{\mathrm{j}\,2\pi k}{N\Delta}\right) = \sum_{n=0}^{N-1} f_n\, \mathrm{e}^{-\mathrm{j}\,2\pi kn/N} = F_k$$

$$\boxed{I_T\left(\frac{\mathrm{j}\,2\pi k}{T}\right) = \frac{\Delta}{N}\left|F_k\right|^2 = \Delta I_N(k)}$$

The discrete-time periodogram is therefore a scaled version of the continuous periodogram.

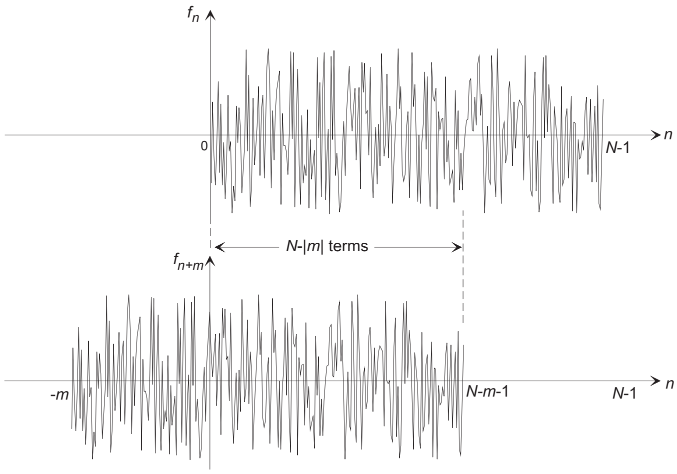
1.1.4 Properties of the Periodogram:

(1) The periodogram is a real, even function. Because it is defined by the Fourier transform of the autocorrelation function (which is a real, even function), the periodogram is also a real, even function.

(2) The periodogram is a biased estimator. Consider the definition of the estimate of the autocorrelation implicitly used above

$$\hat{\phi}_{ff}(m) = \frac{1}{N}\sum_{n=0}^{N-1} f_n f_{n+m}.$$

With a finite length record the overlapping region of the two records $\{f_n\}$ and $\{f_{n+m}\}$ in the summation only includes $N - |m|$ terms:



The estimated autocorrelation function used to compute the priodogram is therefore biased. For the mth lag, the unbiased time-average estimator of the autocorrelation function should therefore be

$$\hat{\phi}'_{ff}(m) = \frac{1}{N-|m|} \sum_{n=0}^{N-1} f_n f_{n+m}, \qquad m = -(N-1), \ldots, N-1$$

which is sometimes known as the *mean-lagged-product*. Then

$$\mathcal{E}\{I_N(k)\} = \sum_{m=-(N-1)}^{N-1} \left(\frac{N-|m|}{N} \right) \hat{\phi}_{ff}(m)\, \mathrm{e}^{-\mathrm{j}\, 2\pi k m/N}$$

The periodogram is therefore a biased estimator of $\Phi_{ff}(\mathrm{j}\,\Omega)$. We note, however, that

$$\lim_{T\to\infty} I_T(\mathrm{j}\,\Omega) = \Phi_{ff}(\mathrm{j}\,\Omega)$$

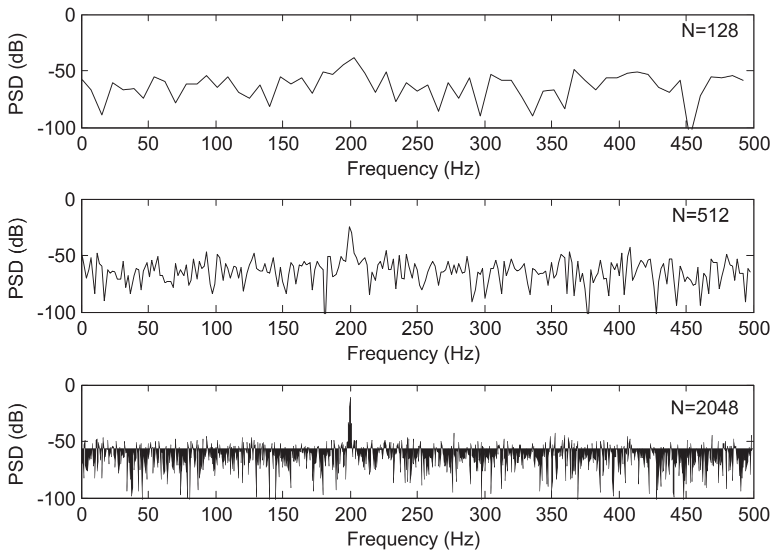
and that $I_T(\mathrm{j}\,\Omega)$ is therefore asymptotically unbiased.

(3) Variance of the periodogram. A somewhat surprising result is that the variance of the of $I_N(k)$ (and $I_T(\mathrm{j}\,\Omega)$) does not decrease significantly as N increases. For large values of the lag m, particularly when $m \approx N$, because there are few values in the sum the variance of the estimate $\hat{\phi}'_{ff}(m)$ becomes large, and although $\hat{\Phi}'_{ff}\left(\mathrm{j}\,\frac{2\pi k}{N}\right)$ is asymptotically unbiased, the variance does not decay to zero as $N \to \infty$. In fact

$$\lim_{N\to\infty} \mathrm{var}[\hat{\Phi}_{ff}(\mathrm{j}\,\Omega)] \approx \Phi^2_{ff}(\mathrm{j}\,\Omega)$$

that is the variance is approximately equal to the square of the true value (see OS&B 10.6, P&M 14.1.2), and is not reduced by taking a longer data record.

As more points are taken into the computation of the periodogram, the apparent spectral resolution is increased, but the reliability of the extra points is marred by the residual variance. The following figure shows periodograms of a 200 Hz sinusoid in noise, sampled with $\Delta = 0.001$ s, and computed with $N = 128, 512,$ and 2048 samples. (Only one side of the periodogram is shown). Notice that the variance does not decrease.



(4) Implicit windowing in the periodogram. We noted above that

$$\mathcal{E}\left\{I_N(\mathrm{j}\,\Omega)\right\} = \sum_{m=-(N-1)}^{N-1} \left(\frac{N-|m|}{N}\right) \phi_{ff}(m)\,\mathrm{e}^{-mj\Omega T}$$

The periodogram is therefore the Fourier transform of the true autocorrelation function multiplied by the Bartlett window function

$$w_N(m) = \frac{N-|m|}{N}, \qquad m = 0, \ldots, N-1.$$

The periodogram is therefore the convolution of the true spectral density and the Fourier transform of the Bartlett window, resulting in a smoothing operation, and the introduction of Gibbs effect ripple adjacent to transitions in the spectrum.
Aside: It is interesting to note that for the Barlett window, $\mathcal{F}\left\{w_N(m)\right\} \geq 0$ and as a result the convolution maintains the requirement that $I_N(\mathrm{j}\,\Omega) \geq 0$.

1.2 Variance Reduction in Periodogram PSD Estimation

There are two common methods of reducing the variance in the periodogram method of PSD estimation (1) the averaging of periodograms, and (2) the smoothing of a single periodogram by windowing the autocorrelation function.

1.2.1 The Bartlett Method: The Averaging of Periodograms

Bartlett proposed the ensemble averaging of basic periodograms as a means of reducing the variance in the estimate - at the expense of spectral resolution, and increased bias. Suppose we have a record $\{f_n\}$ of length N, and we divide the total record into Q smaller blocks, each of length K. Then if we compute the periodogram of each block and designate it $I_K^{(q)}(k)$, $q = 1, \ldots, Q$, the averaged periodogram may be defined as

$$\overline{I}_K(k) = \frac{1}{Q} \sum_{q=1}^{Q} I_K^{(q)}(k)$$

Because each of the smaller blocks contains fewer samples, the spectrum computed through the DFT will have decreased spectral resolution. However, provided each of the Q periodograms are statistically independent, the variance will be reduced by a factor of Q

$$V\left[\overline{I}_K(\mathrm{j}\,\Omega)\right] = \frac{1}{Q} V\left[I_K^{(q)}(\mathrm{j}\,\Omega)\right] \approx \frac{1}{Q} \Phi_{ff}^2(k).$$

1.2.2 The Welch Method: The Averaging of Modified Periodograms

Welch proposed two enhancements to Bartlett's method:

(1) Welch showed that instead of dividing the data sequence into contiguous smaller blocks, it is possible to overlap adjacent blocks by as much as 50% and still maintain statistical independence. The effect is to increase Q, the number of data blocks in the ensemble average and thus effect a further reduction in the variance of $\overline{I}_K(k)$.

(2) Welch's method also applies a window function $w(n)$ to each of the data records before the DFT is computed. A modified periodogram based on the windowed record is then computed

$$\hat{I}_K^{(q)}(k) = \frac{1}{KU} \sum_{n=0}^{K-1} f_n w(n)\, \mathrm{e}^{-\mathrm{j}\, 2\pi nk/Q} \qquad q = 1, \ldots, Q$$

where

$$U = \frac{1}{K} \sum_{n=0}^{K-1} w^2(n)$$

is a factor to compensate for the fact that the windowing operation has reduced the power of the waveform, and allows the estimator to be asymptotically unbiased.

As before, the spectral estimator is taken as the ensemble average of the windowed and overlapped blocks

$$\overline{\hat{I}}_K(k) = \frac{1}{Q} \sum_{q=1}^{Q} \hat{I}_K^{(q)}(k)$$

1.2.3 The Blackman-Tukey Method: Smoothing the Periodogram

Blackman and Tukey proposed that an effective method of variance reduction would be to "smooth" the periodogram estimate with a low-pass filter, and that the smoothing operation could be implemented by applying a suitable *windowing function* in the delay domain of the autocorrelation function, to achieve the desired frequency domain convolution.

An alternative rationale is based on the reliability (variance) of the samples in the autocorrelation function for large lags. As was demonstrated at the start of this lecture, for a fixed length data record, the overlap in the product for computing $\phi_{ff}(m)$ is $N - |m|$, and as $|m| \to N$ the variance becomes large. The windowing operation serves to reduce the contribution of these unreliable estimates in the computation of the PSD.

The Blackman-Tukey estimator is

$$\tilde{I}_N(k) = \sum_{m=-(M-1)}^{M-1} \phi_{ff}(m) w(m)\, \mathrm{e}^{-\mathrm{j}\, 2\pi km/(2M-1)}$$

where the window function is real and symmetric about its mid-point (to ensure that $\tilde{I}_N(k)$ is real). The window length parameter M may be shorter than the data record length N. The Blackman-Tukey estimate is therefore

$$\tilde{I}_N(k) = I_N(k) \otimes W(k)$$

where $\{W(k)\} = \mathrm{DFT}\,\{w(n)\}$.

The choice of window function should be made to ensure that $\tilde{I}_N(\mathrm{j}\,\Omega) > 0$. Many commonly used windows, such as the Hamming and Hann windows, do not have this property, and may result in negative values for the spectral estimates. The triangular Bartlett window does maintain the sign of the estimates.

1.2.4 MATLAB Examples

MATLAB has built-in functions for spectral estimation, in particular the function `spectrum()` is a powerful general function for non-parametric estimation, and the function `pwelch()` can be used for Welch's method.

Notes:

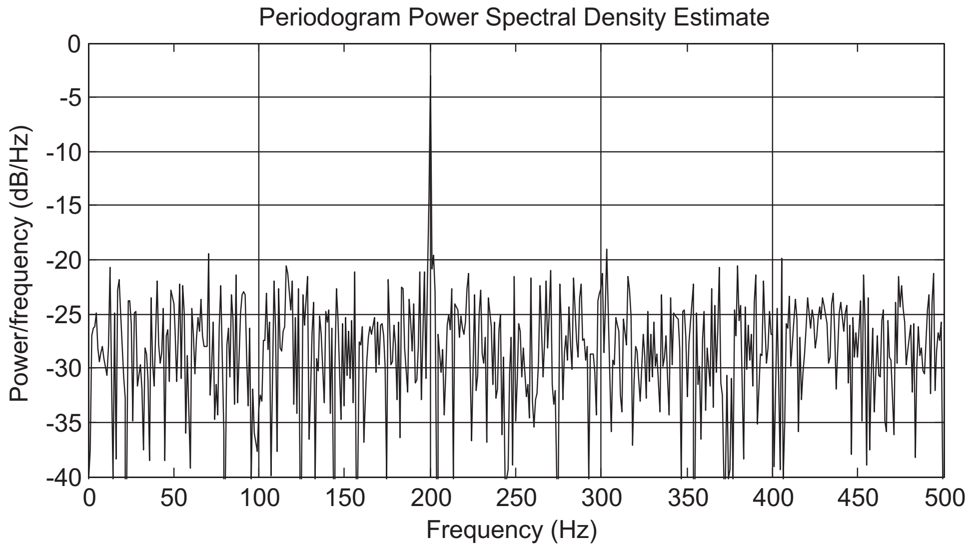
(1) The MATLAB functions can represent a continuous periodogram $I_T(\mathrm{j}\,\Omega)$ by specifying a sampling rate.

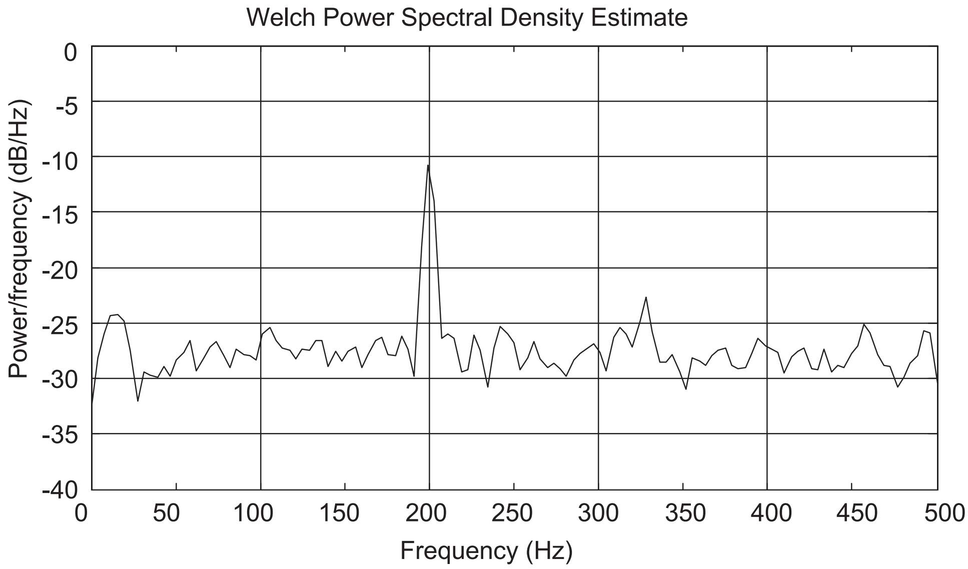
(2) The MATLAB default convention is that if the sample set is real the PSD is computed as a *one-sided* spectrum, that is it is assumed the the power is contained in positive frequencies only. Because of the real, even nature of the periodogram, the one-sided spectrum has values *twice* those of the two-sided spectra. For complex data sets the convention is to compute the two-sided spectra. The defaults can be overuled by optional arguments in the function calls.

The following script was used to display a periodogram and a Welch estimate of a 200 Hz sinusoid in noise. A 1.024 sec. data record, with a sampling rate of 1000 samples/sec. is simulated.

```
% Create the data record.
Fs = 1000;
t = 0:1/Fs:1.024;
f = cos(2*pi*t*200) + randn(size(t));  % A cosine of 200Hz plus noise
% Periodogram
figure(1);
h = spectrum.periodogram;
psd(h,f,'Fs',Fs);
%
% Welch's method.
% Use default of 8 sections, 50% overlap, Hamming window
figure(2);
pwelch(f,128,64,128,Fs);
```

The following two plots were generated.


Periodogram Power Spectral Density Estimate
Power/frequency (dB/Hz)
Frequency (Hz)
0
-5
-10
-15
-20
-25
-30
-35
-40
0
50
100
150
200
250
300
350
400
450
500


Welch Power Spectral Density Estimate
Power/frequency (dB/Hz)
Frequency (Hz)
0
-5
-10
-15
-20
-25
-30
-35
-40
0
50
100
150
200
250
300
350
400
450
500