227-0689-00L System Identification **09 10 2023**

Solution 3: Fourier transform, spectrum and periodogram

Problem 1:

Consider the problem of identifying a linear model for a given plant based on sampled experimental data. Compute the largest sampling time, T_s, and the smallest experiment duration, $t = NT_s$, where N is the number of sample points, required to estimate the transfer function of the plant for at least 256 equally spaced frequencies in the band of $[1\text{Hz}, 100\text{Hz}]$.

Solution

We obtain the largest possible sampling time that fulfills the requirements by setting the Nyquist frequency to be equal to the maximum frequency we are interested in, $f_{\max} = 100\text{Hz}$,

$$f_{\max} = \frac{1}{2T_s}\,. \tag{3.1}$$

To ensure that we have $m = 256$ data samples in the range $[f_{\min}, f_{\max}] = [1\text{Hz}, 100\text{Hz}]$ we require

$$f_0 \leq \frac{f_{\max} - f_{\min}}{m-1}\,, \tag{3.2}$$

where $f_0 := \frac{1}{NT_s}$ is the fundamental frequency specifying the frequency resolution. By combining and rewriting (3.1) and (3.2) we have

$$T_s = \frac{1}{2f_{\max}} = \frac{1}{2 \cdot 100\text{Hz}} = 0.005\text{s}\,,$$
$$N \geq \frac{m-1}{T_s(f_{\max} - f_{\min})} = \frac{2(m-1)f_{\max}}{f_{\max} - f_{\min}} = 515.152\,.$$

We therefore set the sampling time to be $T_s = 5\text{ms}$ and consider experiment data of length $t = NT_s = 516 \cdot 0.005\text{s} = 2.58\text{s}$.

Problem 2:

Let $\{x(k)\}$ be a white Gaussian noise with variance σ_x^2. The periodogram is defined as:

$$\hat{P}_{PER}(\omega_k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-j\omega_k n}\right|^2\,,$$

for $\omega_k = k2\pi/N$ for $k = 0, 1, \ldots, N/2$ (N even).

1. Is the periodogram an unbiased estimator of the PSD?
2. Does the variance of the periodogram go to 0 as $N \to \infty$?
3. Is the periodogram a consistent estimator of the PSD?
4. Assume to have access to M independent datasets of length N. How would you proceed to obtain an estimator with lower variance?

Hint: for a white Gaussian noise:

$$\frac{2\hat{P}_{PER}(\omega_k)}{\sigma_x^2} \sim \chi_2^2 , \qquad k = 1, 2, \ldots, \frac{N}{2} - 1$$

$$\frac{\hat{P}_{PER}(\omega_k)}{\sigma_x^2} \sim \chi_1^2 , \quad k = 0, \frac{N}{2}$$

with χ_ν^2 the *chi-squared distribution* of degree ν.

Solution

Recall that for $y \sim \chi_\nu^2$:

$$E\left[\chi_\nu^2\right] = \nu$$

$$Var\left[\chi_\nu^2\right] = 2\nu$$

Therefore,

$$E\left[\hat{P}_{PER}(\omega_k)\right] = \phi_x(\omega_k) = \sigma_x^2 , \quad k = 0, 1, \ldots, \frac{N}{2}$$

$$Var\left[\hat{P}_{PER}(\omega_k)\right] = \begin{cases} \phi_x^2(\omega_k), & k = 1, 2, \ldots, \frac{N}{2} - 1 \\ 2\,\phi_x^2(\omega_k), & k = 0, \frac{N}{2} \end{cases}$$

If we assume to have access to M independent datasets of length N, we can average M different estimates of the periodogram. The averaged periodogram estimator is

$$\hat{P}_{AVG}(\omega_k) = \frac{1}{M} \sum_{m=0}^{M-1} \left(\frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j\omega_k n} \right|^2 \right) ,$$

for $k \neq 0, \frac{N}{2}$, the variance of the averaged periodogram estimator is

$$\begin{aligned} Var\left[\hat{P}_{AVG}(\omega_k)\right] &= \frac{1}{M^2} \sum_{m=0}^{M-1} \left(Var\left[\hat{P}_{PER}(\omega_k)\right] \right) \\ &= \frac{1}{M^2} M \phi_x^2(\omega_k) \\ &= \frac{1}{M} \phi_x^2(\omega_k) \end{aligned}$$

We can conclude that:

1. The periodogram is an unbiased estimator of the PSD.
2. As $N \to \infty$, the variance of the periodogram does not go to 0. The variance does not decrease with the data length. This is due to the fact that, give N data points, the periodogram tries to estimate about $N/2$ parameters. As more data is available, the the estimation does not improve because also the number of parameters increases proportionally.
3. Therefore, the periodogram is not a consistent estimator of the PSD. Its standard deviation (the square root of the variance) is as large as its expected value, which is the quantity we are trying to estimate.
4. With M independent datasets, the expected value of the averaged periodogram remains unchanged, but the variance is improved by a factor of M.

MATLAB **exercises:**

1. Write a MATLAB function that takes an input vector, $u(k)$, $k = 0, \ldots, N-1$ and returns the vector, $U_N(e^{j\omega_n})$, for

$$\omega_n = \frac{2\pi n}{N}, \quad n = 0, \ldots, N-1.$$

The functional relationship between $u(k)$ and $U_N(e^{j\omega})$ is given by,

$$U_N(e^{j\omega}) = \sum_{k=0}^{N-1} u(k)\,\mathrm{e}^{-j\omega k}, \quad (j = \sqrt{-1}).$$

This looks like an FFT and you may use an `fft` call in your function. See the MATLAB notes section at the end of these exercises for caveats.

2. Write a MATLAB script file (.m file) which performs the following calculations:

   a) Generate $e(k)$, a 1024 point $\mathcal{N}(0,1)$ distributed random sequence.

   b) Calculate and plot (on a log-log scale) the periodogram of $e(k)$. The periodogram is defined as $\frac{1}{N}|E_N(e^{j\omega})|^2$.

   c) Given a discrete-time plant,

$$P(z) = \frac{z+0.5}{0.5 + (z+0.5)(z-0.5)^2},$$

   calculate $w(k)$, the response of $P(z)$ to the input signal, $e(k)$. Assume that the sample time of $P(z)$ is specified as $T = 1$ second.

   d) Calculate the periodogram of $w(k)$.

   e) How is the periodogram of $w(k)$, $\frac{1}{N}|W_N(\mathrm{e}^{j\omega})|^2$, related to $\left|P\left(\mathrm{e}^{j\omega_n}\right)\right|$ (asymptotically) ? Here $\left|P\left(\mathrm{e}^{j\omega_n}\right)\right|$ is the magnitude of the Bode plot of $P(z)$. Provide a plot comparing these quantities. To do this, plot both $\frac{1}{N}\left|W_N(e^{j\omega})\right|^2$ and $\left|P\left(\mathrm{e}^{j\omega_n}\right)\right|^2$ in the same figure. You should also plot $\frac{1}{N}\left|W_N(e^{j\omega})\right|^2 - \left|P\left(\mathrm{e}^{j\omega_n}\right)\right|^2$ to examine the error between the two quantities.

   f) Repeat the above for 2048 and 4096 length sequences. Plot all of your comparisons and see if your assertion in part e) is correct.

MATLAB **notes**

Use the commands `help fft` and `help periodogram` to get more information about these functions. Note that the MATLAB calculation of the DFT (via the FFT algorithm) differs significantly from Ljung's definition. The scale factor, indexing and sign of the exponent are all different. You can also use Ljung's definition but you will have to adjust the answer.

The `periodogram` function in the signal processing toolbox actually estimates the periodogram. It zero pads the data and introduces window functions in its calculation of the estimate. It does not do the calculation given in Ljung for the periodogram.

The functions `rand` and `randn` give different distributions. Make sure that you use the correct one.

Solution hints

1. The definition of the DFT given here coincides with the MATLABfunction `fft`. Different definitions can be found in various literature including [Ljung, 1999]. It is important to be aware of the frequency interval the DFT is being evaluated on.

2. a) Use the MATLAB function `randn` to create a normally distributed random sequence.

   b) Use the function you wrote in 1. and the MATLAB function `abs` to compute the periodogram. Figure 3.1 shows the periodogram of e plotted with the MATLAB function `loglog` for the non-negative frequencies $\omega = 0, \frac{2\pi}{1024}, 2\frac{2\pi}{1024}, \dots, \pi$.

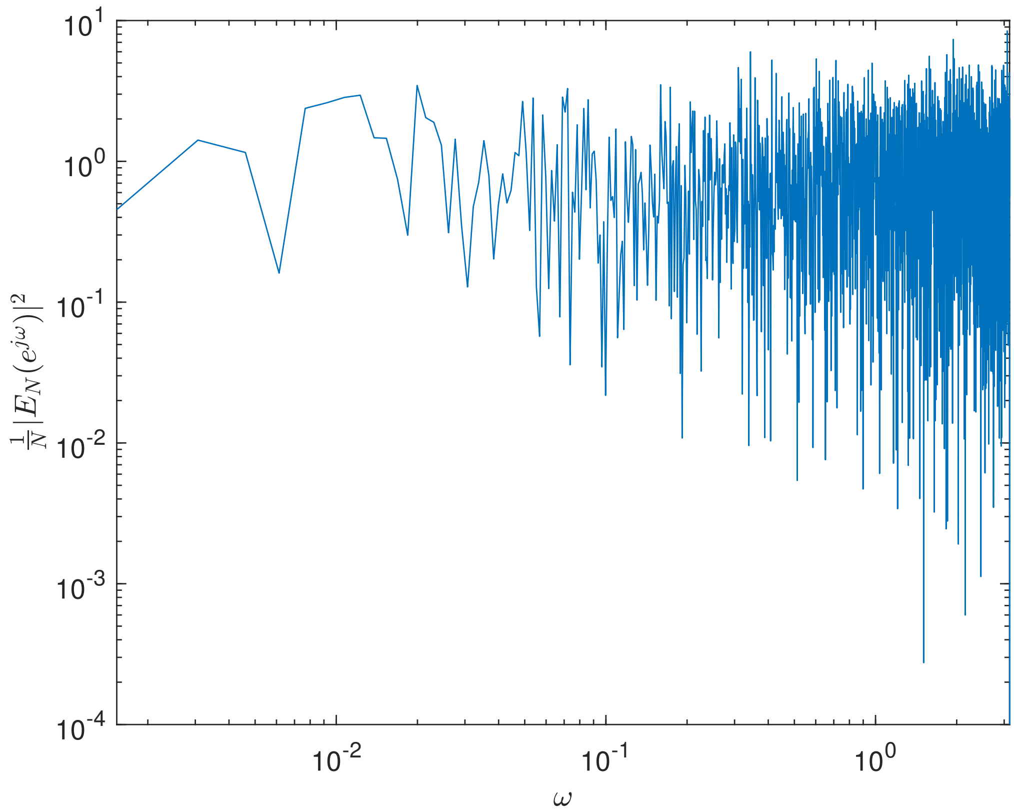

Figure 3.1: Periodogram of e

   c) Define the discrete-time linear time-invariant (LTI) plant in MATLAB with the function `tf`. Define a time vector containing the discrete time steps kT_s, $k = 0, 1, \dots N-1$ and compute the system response to the input signal with the MATLAB function `lsim`.

   d) Calculate the periodogram of the output sequence as done for the input sequence in b).

   e) The periodogram of w is an asymptotically unbiased estimate of the spectrum of w. As the input signal e is a white noise sequence the expected value of periodogram of w converges to the magnitude of the frequency response squared as $N \longrightarrow \infty$, $\lim_{N\longrightarrow\infty} E\left\{\frac{1}{N}|W_N(e^{j\omega})|^2\right\} = |P(e^{j\omega})|^2$. Figure 3.2 compares the periodogram of w with the squared magnitude of the plant transfer function.

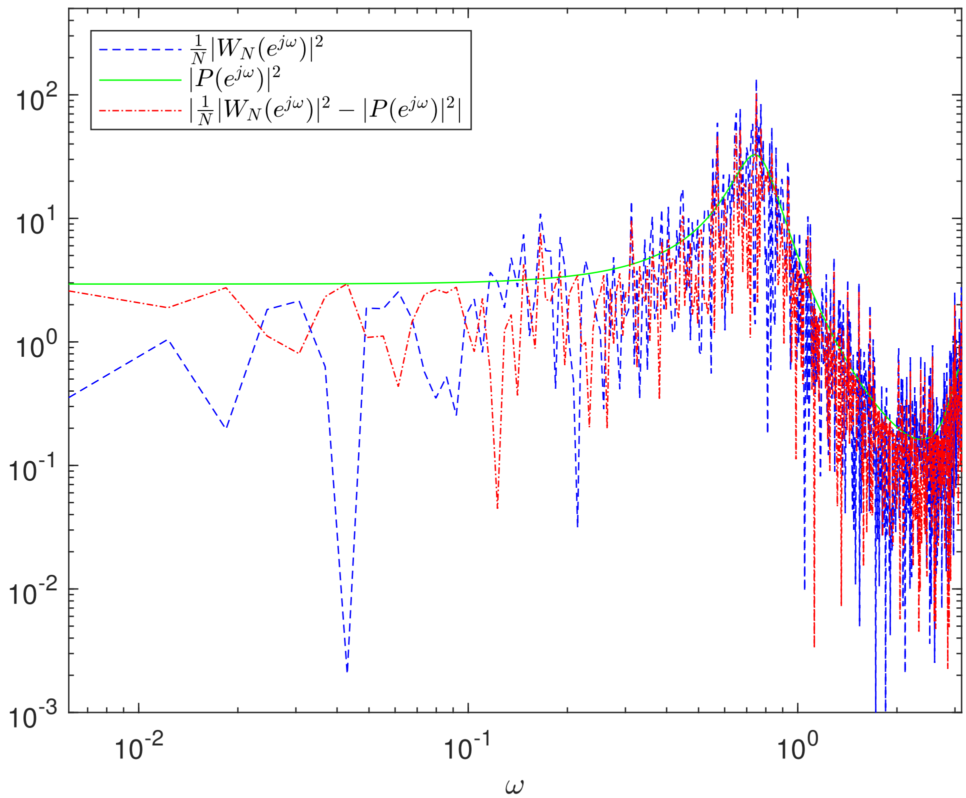

Figure 3.2: Experiment with data length N=1024

f) Figures 3.3 and 3.4 show the results of the above procedure repeated for data lengths 2028 and 4096 respectively. Note that the periodogram still appears jagged for large data lengths as it only converges to the spectrum in expectation.

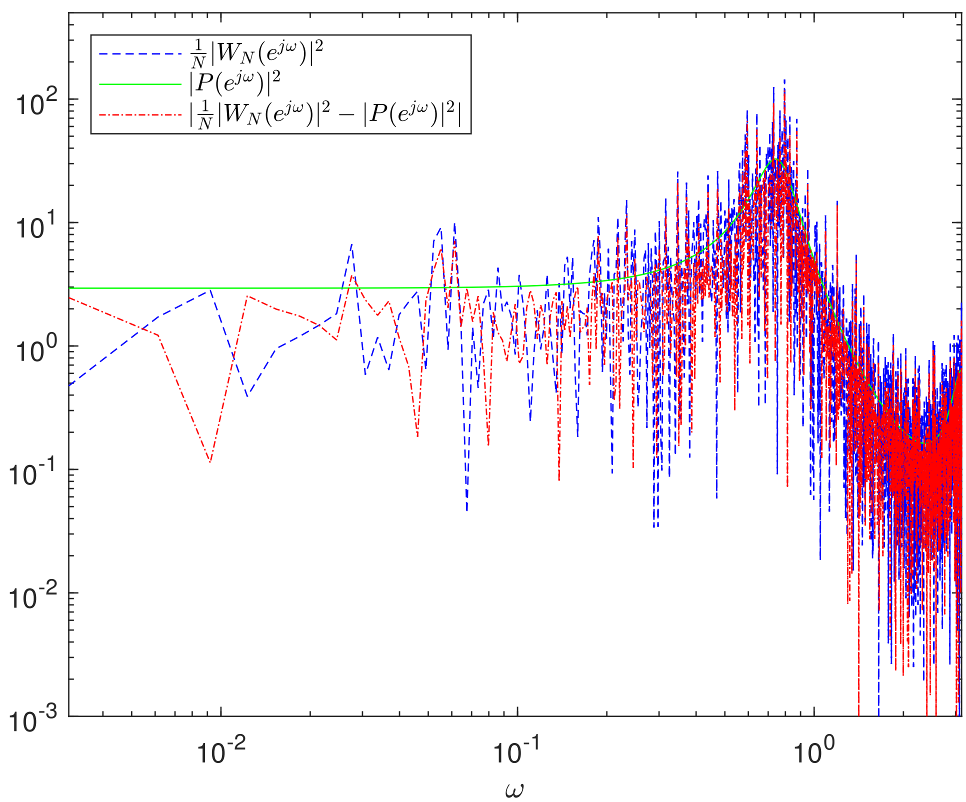

Figure 3.3: Experiment with data length N=2048

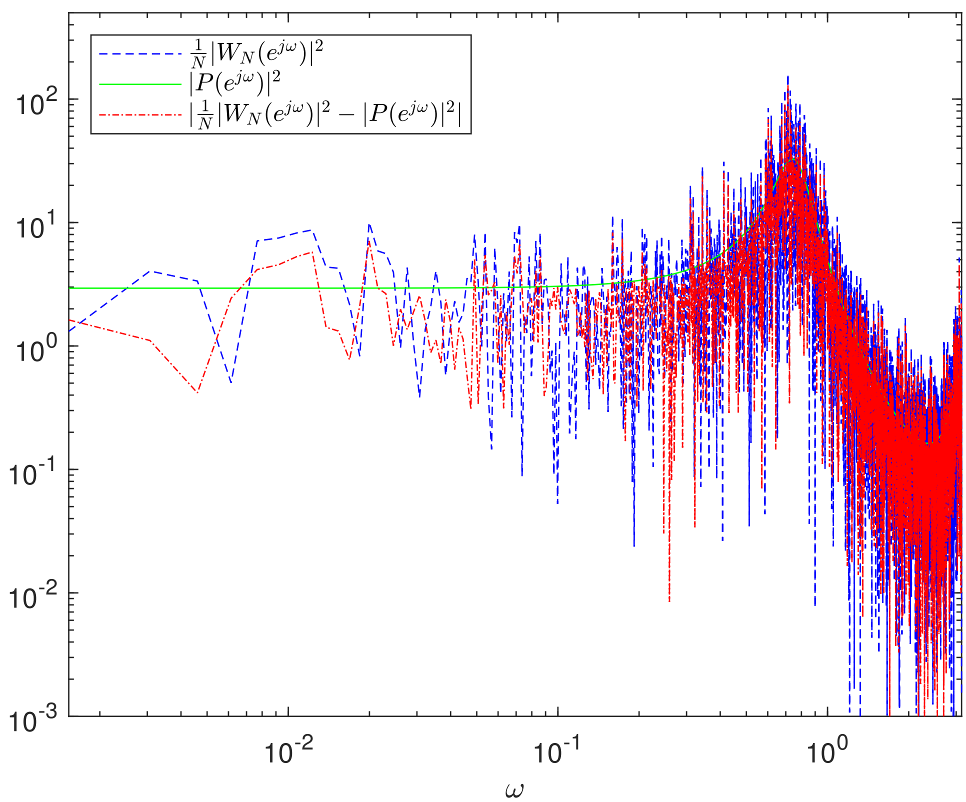

Figure 3.4: Experiment with data length N=4096