Non-parametric Methods

Given

$$u \longrightarrow \boxed{G} \longrightarrow y$$

G unknown but LTI

and

$$\{(u_t, y_t)\}_{t \in T}$$

how can an estimate of a model for G be formed — without making any additional assumptions?

Since G is LTI (and, we assume, causal), it can be described in two ways:

(1) frequency response: $G(\omega)$; $0 \le \omega < \omega$,

(2) impulse response: $G(t, \tau) = G(t - \tau)$.

Reading: Chapter 6 of Ljung

Consider approaches using the frequency domain first:

An obvious estimate of $G(\omega)$ can be formed by finding the Fourier transforms of u and y

Given $T = \{1, \dots, N\}$,

$$U_N(\omega) = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} u(t)\, e^{-j\omega t} \qquad (*)$$

$$Y_N(\omega) = \frac{1}{\sqrt{N}} \sum_{t=1}^{N} y(t)\, e^{-j\omega t}$$

Calculate $$\hat{\hat{G}}_N(e^{j\omega t}) = \frac{Y_N(\omega)}{U_N(\omega)} \quad \text{for } U_N(\omega) \neq 0. \qquad (6.24)$$

$\hat{\hat{G}}_N(e^{j\omega t})$ is undefined for values ω where $U_N(\omega) = 0$.

This is the empirical transfer function estimate (ETFE) in Ljung.

The values of $U_N(\omega)$ and $Y_N(\omega)$ ~~are the~~ at $\omega = 2\pi k/N$, $k = 1, \dots, N$ $(k \in T)$ ~~and~~ are the Discrete Fourier Transforms (DFT) of u_t & y_t, and are equivalent (provide the same information) to $\{(u_t, y_t)\}_{t \in T}$. This can be seen because the time domain sequences can be reconstructed using the inverse DFT:

$$u_t = \frac{1}{\sqrt{N}} \sum_{k=-N/2+1}^{N/2} U_N(2\pi k/N)\, e^{j2\pi k t/N}$$

$$y_t = \frac{1}{\sqrt{N}} \sum_{k=-N/2+1}^{N/2} Y_N(2\pi k/N)\, e^{j2\pi k t/N}$$

and making use of the identities $U_N(\omega + 2\pi) = U_N(\omega)$ and $U_N(-\omega) = \overline{U_N(\omega)}$, and for Y_N.

Other points can be calculated using (*) if needed.

How good is this ETFE ~~$\hat{G}_N(\omega)$~~? $\hat{\hat{G}}_N(e^{j\omega})$?

Assume, for the purpose of error analysis, that

$$y_t = G_0(q)\, u_t + v_t \quad ; \quad G_0(q) \text{ strictly stable}$$

and let $V_N(\omega)$ be the DFT of $\{v_t\}_{t \in T}$.

From Theorem 2.1 (pp 31-33),

$$\hat{\hat{G}}_N(e^{j\omega}) = G_0(e^{j\omega}) + \frac{R_N(\omega)}{U_N(\omega)} + \frac{V_N(\omega)}{U_N(\omega)}$$

where $|R_N(\omega)| \le \frac{K}{\sqrt{N}}$ (K depends upon $\|u\|$ & $\|g\|$)

decreases with $\sqrt{N}$ as N (# of samples) increases.

Assuming $E v_t = 0$ (zero mean),

$$E\, V_N(\omega) = 0 \quad \forall \omega, \quad \text{so}$$

$$E\hat{G}_N(e^{j\omega}) = G_0(e^{j\omega}) + \frac{R_N(\omega)}{U_N(\omega)}$$

With some math (pp 175-6),

$$E\left[V_N(\omega) V_N(-\xi)\right] = \begin{cases} \Phi_v(\omega) + \rho_2(N) & \text{if } \xi = \omega \\ \rho_2(N) & \text{if } |\xi - \omega| = \frac{k2\pi}{N}, k = 1, \dots, N-1 \end{cases}$$

where $|\rho_2(N)| \le 2C/N$.

Applying this to the error in the identified spectrum,

$$\hat{G}_N(e^{j\omega}) - G_0(e^{j\omega})$$

Lemma 6.1 (p. 176):

$$E\hat{G}_N(e^{j\omega}) = G_0(e^{j\omega}) + \frac{\rho_1(N)}{U_N(\omega)}$$

$$|\rho_1(N)| \le \frac{C_1}{\sqrt{N}}$$

$$E\left[\left(\hat{G}_N(e^{j\omega}) - G_0(e^{j\omega})\right)\left(\hat{G}_N(e^{-j\omega}) - G_0(e^{-j\omega})\right)\right]$$

$$= \begin{cases} \dfrac{1}{|U_N(\omega)|^2}\left[\Phi_v(\omega) + \rho_2(N)\right] & \text{if } \xi = \omega \\ \dfrac{\rho_2(N)}{U_N(\omega) U_N(-\xi)} & \text{if } |\xi - \omega| = \frac{2\pi k}{N}, k = 1, \dots, N-1 \end{cases}$$

and $|\rho_2(N)| \le \frac{C_2}{N}$.

So what? Some observations are in order:

1. Look at the effects of $U_N(\omega)$.

Specifically, both bias error and variance increase as $|U_N(\omega)|$ decreases.

This implies that poorly chosen input test signals (or, for many processes, the input signals you are stuck with) can adversely affect spectral estimates.

(a) If the excitation signal has little spectral content in some range of $\omega \in [\omega_1, \omega_2]$ relative to its norm (because C_1 and C_2 depend upon $|u|_\infty$), then the quality of the spectral estimate of G will be relatively poorer for $\omega \in [\omega_1, \omega_2]$ than for other frequencies.

(b) If $U_N(\omega) = 0$ for some ω, no information about G at ω is gained.

(c) The best tactic is to choose u with uniform spectral content over the frequencies of interest in modeling (and zero outside this range, since $\|u\|_\infty = \|U\|_1$.)

For $\hat{G}_N(e^{j\omega})$:

2. Bias error decreases with γ_N as N increases.

3. Error variance has a lower bound ($\neq 0$) as N increases, which is basically the ratio of the process input noise variance to the magnitude squared of the input signal (in frequency domain, evaluated at each frequency).

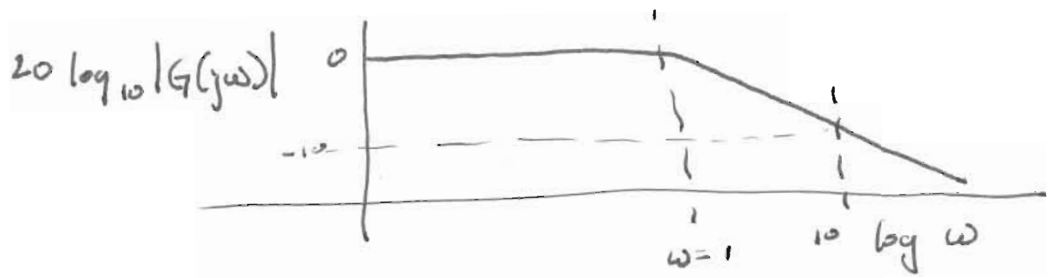
Thus, for ω where an accurate estimate of $G_0(\omega)$ is desired, $|U_N(\omega)|^2 \gg |\Phi_v(\omega)|$ is required.

If we go back to the examples of the first lecture, we can see $\hat{G}_N(e^{j\omega})$ versus N and the residual errors, especially at high frequencies.

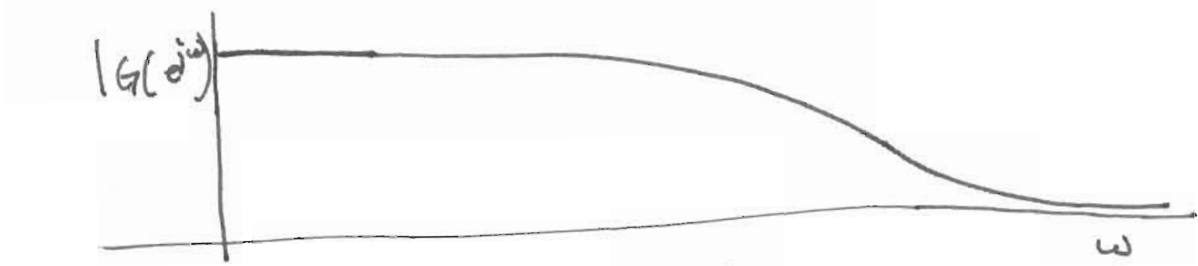
Without making further assumptions about G, one can not to better. However, additional assumptions can be justified.

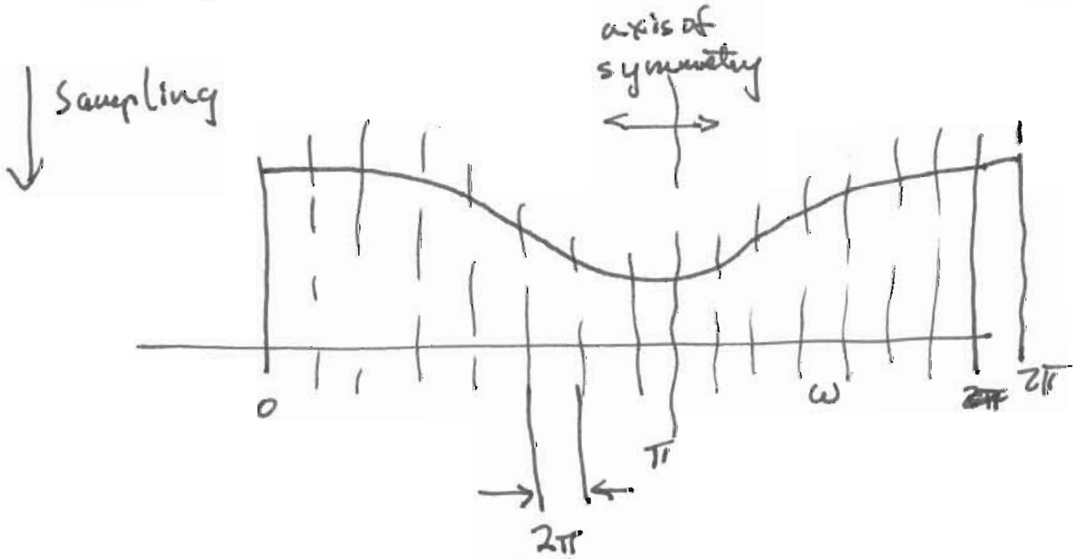
Primary assumption: $G_o(e^{j\omega})$ is typically a continuous and smooth (differentiable) function of ω.

Think of $G(s) = \frac{1}{s+1}$



Without using log/log scales:





If N is sufficiently large (~~too~~ fast sampling rate relative to the time constants of G_o), then one can assume $G_o(e^{j\omega})$ is approximately constant over each of these sub-intervals $\omega \in \left[\frac{2\pi k}{N}, \frac{2\pi(k+1)}{N}\right]$.

If $G_0(e^{j\omega})$ were actually constant over some range of $\omega \in \left[\frac{2\pi k_1}{N}, \frac{2\pi k_2}{N}\right] = [\omega_0 - \Delta\omega, \omega_0 + \Delta\omega]$, given that we know the error variance of the identified $\hat{G}_N(e^{j\omega})$ is bounded by the signal-to-noise ratio (squared), a least-squares problem yields a new (improved?) estimate

$$\hat{G}_N(e^{j\omega_0}) = \frac{\sum_{k=k_1}^{k_2} \alpha_k \hat{\hat{G}}_N(e^{2\pi jk/N})}{\sum_{k=k_1}^{k_2} \alpha_k}$$

where $\alpha_k = \dfrac{|U_N(2\pi k/N)|^2}{\Phi_v(2\pi k/N)}$.

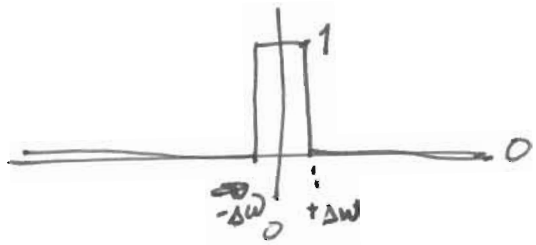
However, ~~is it~~ in fact $G_0(e^{j\omega})$ slowly varies, so it makes sense to prefer information from frequencies near ω_0.

Therefore, a weighting function $W_\gamma(\xi)$ can be introduced. In integral form,

$$\hat{G}_N(e^{j\omega_0}) = \frac{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega_0)\,\alpha(\xi)\,\hat{\hat{G}}_N(e^{j\xi})\,d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega_0)\,\alpha(\xi)\,d\xi}$$

$$\alpha(\xi) = \left[\frac{|U_N(\xi)|^2}{\Phi_v(\xi)}\right]^{-1}$$

For our first estimate, take $W(\xi)$ as



reducing the estimate again to

$$\hat{G}_N(e^{j\omega}) = \frac{\int_{\omega_0-\Delta\omega}^{\omega_0+\Delta\omega} \alpha(\omega)\, \hat{\hat{G}}_N(e^{j\omega})\, d\omega}{\int_{\omega_0-\Delta\omega}^{\omega_0+\Delta\omega} \alpha(\omega)\, d\omega}$$

pp 182 - 187 show definitions of Bartlett, Parzen, and Hamming weighting functions (windows), and their application to a simple system.

Suggested HW:

1. Choose a reasonably simple LTI system (e.g., 2–3rd order & lightly damped, with input and noise.

2. Implement a simulation of the system in Simulink and use the simulation to collect input/output data sets for a variety of inputs:

   (a) white noise (or random walk)

   (b) output of a 1st order system driven by white noise

   (c) Random transitions between -1 an 1 (fig 6.2, p.185 is an example)

   (d) Square wave

3. For varying sampling rates and values of N, implement software to calculate $\hat{\hat{G}}_N(e^{j\omega})$ and $\hat{G}_N(e^{j\omega})$ using a Parzen window.

4. Compare your results. How do they depend upon input, noise, sampling rate, and N?