2.160 System Identification, Estimation, and Learning
Lecture Notes No. 18
April 19, 2006

11 Informative Data Sets and Consistency

11.1 Informative Data Sets

Predictor: $\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + \left[1 - H^{-1}(q)\right]y(t)$

$$\hat{y}(t|t-1) = \left[W_u(q), \quad W_y(q)\right]\begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = W(q)z(t) \tag{1}$$

<u>Definition1</u> Two models $W_1(q)$ and $W_2(q)$ are equal if frequency functions

$$W_1(e^{i\omega}) = W_2(e^{i\omega}) \tag{2}$$

for almost all $\omega \quad -\pi \le \omega \le \pi$

<u>Definition2</u> A quasi-stationary data set Z^∞ is informative enough with respect to model structure M if, for any two models in M

$\hat{y}_1(t|\theta_1) = W_1(q)z(t)$ and $\hat{y}_2(t|\theta_2) = W_2(q)z(t)$

Condition

$$\overline{E}[(\hat{y}_1(t|\theta_1) - \hat{y}_2(t|\theta_2))^2] = 0 \tag{3}$$

implies

$$W_1(e^{i\omega}) = W_2(e^{i\omega}) \tag{4}$$

for almost all $\omega \quad -\pi \le \omega \le \pi$

Let us characterize a quasi-stationary data set Z^∞ by power spectrum $\Phi_v(\omega)$ (Spectrum Matrix):

$$\Phi_z(\omega) = \begin{bmatrix} \Phi_u(\omega) & \Phi_{uy}(\omega) \\ \Phi_{yu}(\omega) & \Phi_y(\omega) \end{bmatrix} \in R^{2\times 2} \tag{5}$$

Theorem 1 A quasi-stationary data set Z^∞ is <u>informative</u> if the spectrum matrix for $z(t) = (u(t), y(t))^T$ is strictly positive definite for almost all ω.

Proof

$\hat{y}_1(t|\theta_1) - \hat{y}_2(t|\theta_2) = \left[W_1(q) - W_2(q)\right]z(t)$

Using eq.11 of Lecture Note 17, (3) can give by

$$\overline{E}\left[(W_1 - W_2)z(t)\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[W_1(e^{i\omega}) - W_2(e^{i\omega})\right]^T \Phi_z(\omega)\left[W_1(e^{i\omega}) - W_2(e^{i\omega})\right]d\omega = 0 \tag{6}$$

Since $\Phi_z(\omega)$ is strictly positive definite for almost all $\omega \quad -\pi \le \omega \le \pi$, the above integral becomes zero only when the vector of the quadratic form, W1-W2, is zero for almost all ω. Namely,

$W_1(e^{i\omega}) \equiv W_2(e^{i\omega})$ for almost all $\omega \quad -\pi \le \omega \le \pi$

Remark: Theorem 1 applies to an arbitrary linear model set. As long as the spectrum matrix $\Phi_z(\omega)$ is strictly positive definite, the data set can distinguish any two linear mocels, regardless of model structure, ARX,OE etc. Also this applies to closed-loop systems, where $\Phi_{uy}(\omega) \neq 0$.

11.2 Consistency of Prediction Error Based Estimate

The prediction-error estimate is defined as

$$\hat{\theta}_N = \arg\min_{\theta \in D_M} V_N(\theta, Z^N) \tag{7}$$

$$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N}\frac{1}{2}\varepsilon^2(t,\theta) \tag{8}$$

The original problem is to find $\hat{\theta}$ that minimizes the expected (ensemble mean) squared prediction error:

$$\overline{V}(\theta, Z^N) = \overline{E}\left[\frac{1}{2}\varepsilon^2(t,\theta)\right] \tag{9}$$

However, the erogicity:

$$\lim_{N\to\infty} V_N(\theta, Z^N) = \overline{V}(\theta) \tag{10}$$

Holds if, (the following conditions are for mathematical rigor)

1) the model structure is uniformly (in θ) stable and linear,
2) {y(t),u(t)} are jointly quasi-stationary,
3) y(t) and u(t) are generated with uniformly stable filters, and
4) y(t) and u(t) are driven by
   - bounded, deterministic inputs, and/or
   - independent random variables with zero means bounded moments of

True System

Let us assume that the actual data are generated by the following "true system"

$$S: \quad y(t) = G_0(q)u(t) + H_0(q)e_0(t) \tag{11}$$

Where $H_o(q)$ is inversely stable(inverse is also stable) and monic, and $\{e_o(t)\}$ is a sequence of random variables with zero mean values, variances λ_0 and bounded moments of order 4+δ.

When the true system is involved in a model structure

$$M: \quad \{G(q,\theta), H(q,\theta) | \theta \in D_M\} \tag{12}$$

The following set of model parameters equal to the true system is not empty:

$$D_T(S,M) = \{\theta \in D_M \,|\, G(e^{i\omega},\theta) = G_0(e^{i\omega},\theta), H(e^{i\omega},\theta) = H_0(e^{i\omega},\theta); -\pi \le \omega < \pi\} \quad (13)$$

Theorem 2 Let M be a linear, uniformly stable model structure containing a true system $S \in M$. If a quasi-stationary data set Z^{∞} is informative enough with respect to M, then the prediction errors estimate is <u>consistent</u>:

$$\arg\min_{\theta\in D_M} \overline{V}(\theta, Z^N) = \lim_{N\to\infty} \arg\min_{\theta\in D_M} V_N(\theta, Z^N) \in D_T(S,M) \qquad (14)$$

If, in addition, the parameter of the true system is unique, $D_T(S,M) = \{\theta_0\}$,then

$$\lim_{N\to\infty} \arg\min_{\theta\in D_M} V_N(\theta, Z^N) = \theta_0; \qquad (15)$$

Proof Consider the difference between $\overline{V}(\theta) = \overline{V}(\theta_0)$ for arbitrary $\theta \in D_M$ and the true system's parameter vector θ_0,

$$\begin{aligned} \overline{V}(\theta) - \overline{V}(\theta_0) &= \overline{E}\left[\frac{1}{2}\varepsilon^2(t,\theta)\right] - \overline{E}\left[\frac{1}{2}\varepsilon^2(t,\theta_0)\right] \\ &= \frac{1}{2}\overline{E}\left[\left(\varepsilon^2(t,\theta) - \varepsilon^2(t,\theta_0)\right)^2\right] + \underbrace{\overline{E}\left[\left(\varepsilon(t,\theta) - \varepsilon(t,\theta_0)\right)\cdot\varepsilon(t,\theta_0)\right]}_{\text{(A)}} \end{aligned} \qquad (16)$$

Compute $\varepsilon(t,\theta_0)$ using the true system assumption (11)

$$\varepsilon(t,\theta_0) = y(t) - \hat{y}(t|\theta) = -H_0^{-1}(q)G_0(q)u(t) + H_0^{-1}(q)y(t) = e_0(t) \qquad (17)$$

There $\varepsilon(t,\theta_0) = e_0(t)$ is an independent random variable of zero mean values. In (A) is given by $\varepsilon(t,\theta\ ) - \varepsilon(t,\theta_0)$ is given by

$$\varepsilon(t,\theta\ ) - \varepsilon(t,\theta_0) = \hat{y}(t|\theta_0) - \hat{y}(t|\theta) \qquad (18)$$

which depends on Z^{t-1}, the input-output data upto t-1.
Therefore, it is uncorrelated with e(t), i.e. (A)=0.

$$\overline{V}(\theta) - \overline{V}(\theta_0) = \frac{1}{2}\overline{E}\left[\left(\hat{y}(t|\theta) - \hat{y}(t|\theta_0)\right)^2\right] \qquad (19)$$

From Theorem 1, since Z^{∞} is informative enough, as long as the two models corresponding to θ and θ_0 are different $\overline{E}\left[\left(\hat{y}(t|\theta) - \hat{y}(t|\theta_0)\right)^2\right] > 0$. This means that

$$\overline{V}(\theta) > \overline{V}(\theta_0) \text{ for all } \theta \ne \theta_0 \qquad (20)$$

11.3 Frequency Domain Analysis of Consistency

Using eq.(11), the mean prediction error can be written as

$$\overline{V}(\theta)=\overline{E}\left[\frac{1}{2}\varepsilon^2(t,\theta)\right]=\frac{1}{4\pi}\int_{-\pi}^{\pi}\Phi_{\varepsilon}(\omega,\theta)d\omega \tag{21}$$

where $\Phi_{\varepsilon}(\omega,\theta)$ is the power spectrum of the prediction error $\{\varepsilon(t,\theta)\}$. Based on the true system description (11)

$$\begin{aligned}\varepsilon(t,\theta)&=H_{\theta}^{-1}\left[y(t)-G_{\theta}u(t)\right]=H_{\theta}^{-1}\left[(G_0-G_{\theta})u(t)+H_0e_0(t)\right]\\&=H_{\theta}^{-1}\left[(G_0-G_{\theta})u(t)+(H_0-H_{\theta})e_0(t)\right]+e_0(t)\end{aligned} \tag{22}$$

$$\Phi_{\varepsilon}(\omega,\theta)=\frac{|G_0-G_{\theta}|^2}{|H_{\theta}|^2}\Phi_u(\omega)+\frac{|H_0-H_{\theta}|^2}{|H_{\theta}|^2}\lambda_0+\lambda_0 \tag{23}$$

For an open-loop system with $\Phi_{eu}(\omega)=0$

It follows directly from (21) and (23) that, if there exist the parameter vector such that $G_{\theta_0}=G_0$ and $H_{\theta_0}-H_0$, then such θ_0 minimizes $\overline{V}(\theta)$, the equivalent result to Theorem 2.

Consider a case that noise model $H(q,\theta)$ has been known as fixed : $H(q,\theta)=H^*(q)$.
The minimization of $\overline{V}(\theta)$ is then reduced to

$$\hat{\theta}=\arg\min_{\theta}\int\left|G_0(e^{i\omega})-G(e^{i\omega},\theta)\right|^2\cdot\frac{\Phi_u(\omega)}{\left|H^*(e^{i\omega})\right|^2}d\omega \tag{24}$$

Remarks:

- The model $G(q,\theta)$ is pushed towards the true system $G_o(q)$ in such a way that the <u>weighted</u> mean squared difference in the frequency domain be minimized.
- The weight, $\dfrac{\Phi_u(\omega)}{\left|H^*(e^{i\omega})\right|^2}$, is the ratio of the input power spectrum to the noise power spectrum(if the variance of $e_o(t)$ is unity). In other words, it is a signal-to-noise ratio.