System Identification

6.435

SET 8

- Convergence and Consistency
- Informative Data (relation to p.e.)
- Convergence to the true parameters
  (role of identifiability)

Munther A. Dahleh

Convergence and Consistency

Estimator

$$Z^N \longrightarrow \widehat{\theta}_N \in D_m$$

Question: Given certain properties of Z^N, and a particular method for arriving to $\widehat{\theta}_N$, what properties does $\widehat{\theta}_N$ have?

- Does $\widehat{\theta}_N \longrightarrow \theta^*$?
- Does $\widehat{\theta}_N \longrightarrow$ "set" ?

Ergodicity Result

<u>Theorem (Ljung)</u>

Let $\{G_\theta(q), \theta \in D_\theta\}$ be a uniformly stable family of filters, ω_θ is a family of deterministic signals such that

$$|\omega_\theta(t)| \leq C_W \quad \forall \quad \theta \in D_\theta$$

Let the signal $s_\theta(t)$ be defined (for each θ) as

$$s_\theta(t) = G_\theta(q)v(t) + \omega_\theta(t)$$

where $v(t)$ is a quasi-stationary signal generated by

$$v(t) = H_t(q)e(t)$$

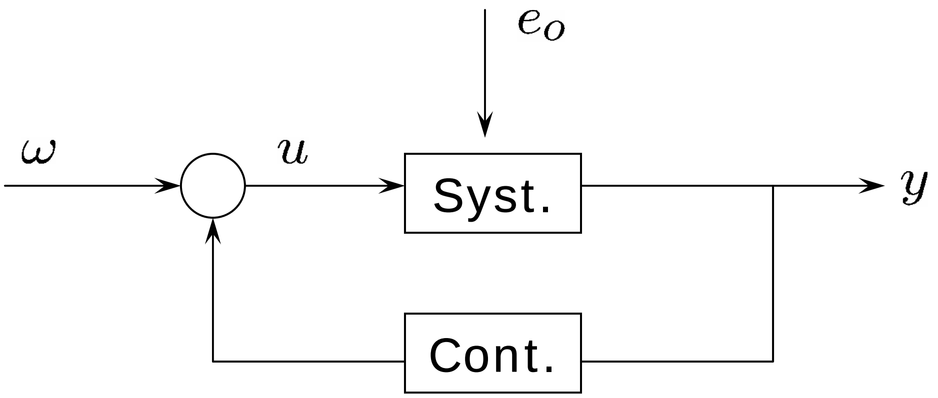
where H_t is a uniformly stable family of filters, and e is white with $E\left(ee^T(t)\right) = \Lambda$. Then,

$$\sup_{\theta \in D_\theta} \left\| \frac{1}{N} \sum_{t=1}^{N} s_\theta(t) s_\theta^T(t) - E s_\theta(t) s_\theta^T(t) \right\| \to 0$$

as $N \to \infty$ w.p.1

Assumptions

1. Data is generated in either open loop or close loop:



$$y = H_1 \omega + H_2 e_o$$
$$u = H_3 \omega + H_4 e_o$$

ω - exogenous input

e_o - noise

$$\begin{cases} \omega(t) & \text{- deterministic, bounded} \\ e_o(t) & \text{- white signal, and bounded} \\ & \text{moments of order higher than 4.} \end{cases}$$

$$\begin{cases} H_i \text{ are stable transfer functions.} \\ \boldsymbol{y} \text{ \& } \boldsymbol{u} \text{ are jointly quasi-stationary} \end{cases}$$

Remark: We view e_o as a stochastic signal & everything else as deterministic. Hence, $E(\cdot)$ is with respect to e_o.

True system:

$$\delta: \quad y(t) = G_o(q)u + H_o(q)u$$

e_o white, bdd 4 moments.

Model structure:

$$m: \quad \{G(q,\theta), H(q,\theta) | \theta \in D_m\}$$

Feasible set:

$$D_T(\xi, m): \left\{\theta \in D_m | G\left(e^{i\omega}, \theta\right) = G_o\left(e^{i\omega}\right), \right.$$
$$\left. H\left(e^{i\omega}, \theta\right) = H_o\left(e^{i\omega}\right); -\pi \leq \omega \leq \pi\right\}$$

Input Choice:

$$u = -F(q)y + \omega$$

Such that F stabilizes G_o. Then previous assumption on Data holds.

Informative Data

- Informativity is a notion that will allow us to distinguish between different models in a structure.
- $Z^N = \left(u(t) \;\; , \;\; y(t) \;\; ; \;\; t \leq N-1 \right)$

  $Z^\infty = \left(u(t) \;\; , \;\; y(t) \;\; ; \;\; \forall \quad t \right)$
- <u>Def</u>: A quasi-stationary data set Z^∞ is informative enough with respect to a model set m^* if for any $W_1, W_2 \in m^*$,

  $$\bar{E}((W_1(q) - W_2(q))Z(t))^2 = 0 \Rightarrow W_1\left(e^{i\omega}\right) = W_2\left(e^{i\omega}\right) \quad \text{a.e.}$$
- Detail $\quad (W_1(q) - W_2(q))Z(t) = \Delta W_u(q)u + \Delta W_y(q)y$
- <u>Def</u>: Z^∞ is informative if it is informative enough with respect to all LTI models.

- Thm: Z^{∞} is informative if the spectrum of $z = \begin{bmatrix} u & y \end{bmatrix}^T$ is strictly positive definite $\forall \quad \omega$

- Proof: Let $\tilde{W} = W_1 - W_2$

- $\bar{E}(\tilde{W}(q)z)^2 = 0 \quad \Leftrightarrow \quad \int_{-\pi}^{\pi} \tilde{W}\left(e^{i\omega}\right)\Phi_z(\omega)\tilde{W}^T\left(e^{-i\omega}\right)d\omega = 0$

$$\Updownarrow \qquad\qquad \Updownarrow$$

$$\tilde{W}(q) = 0 \quad \Leftrightarrow \quad \tilde{W}\left(e^{i\omega}\right)\Phi_z(\omega)\tilde{W}^T\left(e^{-i\omega}\right) = 0 \quad \text{a.e.}$$

$$\Rightarrow \quad \Phi_z(\omega) > 0 \quad \text{a.e.}$$

- What is the relation between informativity of data & persistence of excitation of an input? Recall, it is the input that you can choose!

Informativity vs. Persistence of Excitation

- Thm: Let $m = \{G(q,\theta), H(q,\theta) | \theta \in D_m\}$ and assume that

  $$G(q,\theta) = \frac{B(q)}{F(q)}$$

  where B, F are polynomials of order n_b, n_f respectively.
  If u is p.e. of order $n_b + n_f$, then the data record $[u, y]$ is informative with respect to m.

- Proof: We claim that data is informative if for any
  $\Delta G = G_1 - G_2, G_i \in m,$

  $$\left|\Delta G\left(e^{i\omega}\right)\right| \Phi_u\left(e^{i\omega}\right) = 0 \Leftrightarrow \Delta G = 0$$

If this holds, then u is p.e. of order $n_b + n_f, \Phi_u(\omega) > 0$ for at least $n_b + n_f$ frequencies.

- Proof of Claim:

$$W_1 z - W_2 z = \hat{y}_1 - \hat{y}_2 = (y - \hat{y}_2) - (y - \hat{y}_1) = \varepsilon_2 - \varepsilon_1$$

$$\varepsilon_1 = H_1^{-1}(y - G_1 u)$$

$$\varepsilon_2 = H_2^{-1}(y - G_2 u) \qquad \text{where} \quad G_i, H_i \in m$$

also note that

$$\varepsilon_2 = H_2^{-1}(G_o u - G_2 u + H_o e); \qquad G_o, H_o \text{ are the true models.}$$

$$\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$$

$$= H_2^{-1}(y - G_2 u) - H_1^{-1}(y - G_1 u)$$

$$= H_1^{-1}\left[G_1 u - y + \frac{H_1}{H_2} y - \frac{H_1}{H_2} G_2\right]$$

$$= H_1^{-1}\left[(G_1 - G_2)u + \frac{(H_1 - H_2)}{H_2} y - \frac{(H_1 - H_2)}{H_2} G_2 u\right]$$

$$= H_1^{-1}[\Delta G u + \Delta H \varepsilon_2]$$

$$\bar{E}(\Delta\varepsilon)^2 = \bar{E}\left(\frac{1}{H_1}\left[\Delta G u + \frac{\Delta H}{H_2}(G_o - G_2)u + \frac{\Delta H}{H_2} H_o e\right]\right)^2$$

$$= \int_{-\pi}^{\pi} \frac{1}{|H_1|^2}\left|\Delta G + \frac{\Delta H}{H_2}(G_o - G_2)\right|^2 \Phi_u(\omega) d\omega$$

$$+ \int_{-\pi}^{\pi} \frac{|\Delta H|^2}{|H_1|^2 |H_2|^2} |H_o|^2 \lambda^2 d\omega = 0$$

$\Rightarrow$ Both integrals = 0

$$\Rightarrow \frac{|\Delta H|^2}{|H_1|^2|H_2|^2}|H_o|^2\lambda^2 = 0 \qquad \text{But } H_o \neq 0$$

$$\Rightarrow |\Delta H|^2 = 0 \quad \Rightarrow \quad H_1 = H_2 \quad \text{a.e.}$$

(Comment: Richness of noise guarantees $H_1 = H_2$).

$$\Rightarrow \frac{1}{{H_1}^2}|\Delta G|^2\Phi_u(\omega) = 0 \quad \text{a.e.}$$

$$\text{i.e.} \quad \bar{E}(\Delta\varepsilon)^2 = 0 \quad \Leftrightarrow \quad \begin{cases} H_1 = H_2\left(e^{i\omega}\right) \\ |\Delta G|^2\Phi_u(\omega) = 0 \end{cases}$$

Informativity $\Leftrightarrow$ Persistence of excitation w.r.t. to $\boldsymbol{G}$.

Assumptions

Def: m is uniformly stable if the family of filters

$$\left\{W(q,\theta),\Psi(q,\theta),\frac{d}{d\theta}\Psi(q,\theta)\right\}$$ is uniformly stable.

More assumptions:

1) Model structure is uniformly stable.

2) $V_N''(\theta) \triangleq$ The Hessian is non-singular, at least locally around $\min_{\theta} V(\theta)$.

3) Data is informative.

Analysis of Prediction Error Methods

$$\hat{\theta}_N = \underset{\theta \in D_m}{\text{argmin}} \; V_N\left(\theta, Z^N\right)$$

Quadratic objective $V_N\left(\theta, Z^N\right) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t,\theta)$

$$\varepsilon(t,\theta) = y(t) - Wz = (1 - W_y)y + (-W_u)u$$

It follows that:

$\varepsilon(t,\theta) = H_5(q,\theta)w + H_6(q,\theta)e_o, \quad H_5, H_6$ are uniformly stable.

This is a result of uniform stability of $\boldsymbol{W}$, and the fact that $\boldsymbol{y}$, $\boldsymbol{u}$ are generated by $\boldsymbol{W}$ & e_o through stable filters.

Lemma: Let the assumptions on

- Data generation
- Uniform stability of model structure

hold. Then,

$$\sup \left\| \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t,\theta) - E\varepsilon^2(t,\theta) \right\| \to 0 \quad \text{as } N \to \infty \quad \text{w.p.1}$$

Proof: follows immediately from the basic ergodicity theorem.

Let

$$\bar{V}(\theta) = \bar{E}\frac{1}{2}\varepsilon^2(t,\theta)$$

$$D_C = \underset{\theta \in D_m}{\text{argmin}}\ \bar{V}(\theta) = \left\{ \theta' \middle| \theta' \in D_m, \bar{V}\left(\theta'\right) = \min_{\theta}\ \bar{V}(\theta) \right\}$$

(all possible solutions).

Thm: Under the assumptions of the lemma,

$$\hat{\theta}_N \to D_C \quad \text{w.p.1} \quad \text{as } N \to \infty$$

$$\left(\inf_{\theta \in D_C} \left| \hat{\theta}_N - \theta \right| \to 0 \quad \text{w.p.1} \quad \text{as } N \to \infty \right)$$

Proof: From previous lemma, $V_N\left(\theta, Z^N\right)$ converges to $\bar{V}(\theta)$ uniformly on D_m. The result follows immediately from this.

Example

δ : $y + a_o y(t-1) = b_o u(t-1) + e_o(t) c_o e_o(t-1)$ u, e_o are white.

Model structure $\boldsymbol{m}$: $\widehat{y} = -ay(t-1) + bu(t-1)$

$$\theta = \begin{pmatrix} a \\ b \end{pmatrix}$$

Previously, we have computed all the expectations for

$$\widehat{\theta}_N = \text{argmin } \frac{1}{N} \sum_{t=1}^{N} |y - \widehat{y}(t)|^2 \quad \text{(Least Squares)}$$

$$\bar{V}(\theta) = \bar{E}\varepsilon^2(t,\theta) = \bar{E}(y(t)+ay(t-1)-bu(t-1))^2$$

$$= r_o(1 + a^2 - 2aa_o) + b^2 - 2bb_o + 2ac_o$$

$$\& \quad r_o = Ey^2 = \frac{{b_o}^2 + c_o(c_o - a_o) - a_o c_o + 1}{1 - {a_o}^2}$$

The values $\begin{cases} \hat{a} = a_o - \frac{c_o}{r_o} \\ \hat{b} = b_o \end{cases}$ minimize $\bar{V}(\theta)$

$$\bar{V}(\hat{\theta}) = 1 + {c_o}^2 - \frac{{c_o}^2}{r_o}$$

which is smaller than the variance for θ_o

$$\bar{V}(\hat{\theta}_o) = 1 + {c_o}^2.$$

Of course, the estimate depends on the input.

Example

$$\delta: \quad y(t) = b_o u(t-1) + e_o(t)$$

$$u(t) = d_o u(t-1) + \omega(t) \qquad e_o, \omega \text{ are indep.}$$

Model structure: $\hat{y}(t,\theta) = bu(t-2) \quad , \quad \theta = b.$

$$E(y(t) - bu(t-2))^2 = E(b_o u(t-1) - bu(t-2))^2 + E{e_o}^2$$

$$= E((b_o d_o - b)u(t-2) + b_o \omega(t-1))^2 + 1$$

$$= \frac{(b_o d_o - b)^2}{1 - {d_o}^2} + {b_o}^2 + 1.$$

$$\underset{b}{\text{argmin}}\ \bar{V}(\theta) = b_o d_o$$

equiv. $\hat{b}_N \to b_o d_o$ as $N \to \infty$

$$\bar{V}(\hat{\theta}_N = \hat{b}_N) = 1 + b_o{}^2$$

$$\Rightarrow \quad \hat{y}(t) = b_o d_o u(t-2)$$

If u is white, i.e., $d_o = 0$, then

$$\hat{y}(t) = 0$$

Consistency & Convergence

Theorem: Assume that

a) Z^{∞} is informative enough w.r.to m

b) $\delta \in m$ (equiv. $D_T \neq \phi$)

Then,

1) $$D_C \hat{=} \{\theta = \underset{\theta \in D_m}{\text{argmin}} \; \bar{V}(\theta)\} = D_T$$

2) If the model structure is globally identifiable at θ_o, then
$$D_C = D_T = \{\theta_o\}$$

3) $$G\left(e^{i\omega}, \hat{\theta}_N\right) \rightarrow G_o\left(e^{i\omega}\right) \quad , \quad H\left(e^{i\omega}, \hat{\theta}_N\right) \rightarrow H_o\left(e^{i\omega}\right)$$

w.p.1 as $N \rightarrow \infty$

Proof: To gain intuition we will prove the open loop case 1st and then the closed loop.

Case I Open Loop experiment. This implies that $u(t)$ & $e(t)$ are indep.

Recall: $\varepsilon(t,\theta) = y - \hat{y}(t,\theta)$

$$= H^{-1}(q,\theta)[y - G(q,\theta)u]$$

$$= H^{-1}(q,\theta)[G - G(q,\theta)]u + H^{-1}(q,\theta)H(q)$$

$$\bar{E}\left(\varepsilon^2(t,\theta)\right) = \int_{-\pi}^{\pi} \left|\frac{1}{H\left(e^{i\omega},\theta\right)}\right|^2 \left|G\left(e^{i\omega}\right) - G\left(e^{i\omega},\theta\right)\right|^2 \Phi_u(\omega)d\omega$$

$$+ \int \left|1 - H^{-1}\left(e^{i\omega},\theta\right) H\left(e^{i\omega}\right)\right|^2 \lambda^2 d\omega + \lambda^2$$

$$\left(H^{-1}(q,\theta)H(q) = e(t) + \left(1 - H^{-1}(q,\theta)H(q)\right)e\right)$$

$$\Rightarrow \quad \bar{E}\left(\varepsilon^2(t,\theta)\right) \geq \lambda^2$$

Equality holds if $1 - H^{-1}\left(e^{i\omega},\theta\right)H\left(e^{i\omega}\right) = 0$

$$\& \quad \left|G\left(e^{i\omega}\right) - G\left(e^{i\omega},\theta\right)\right|^2 \Phi_u(\omega) = 0$$

If u is p.e. of order $n_b + n_f - 1 \Rightarrow G\left(e^{i\omega,\theta}\right) = G\left(e^{i\omega}\right)$

$$\Rightarrow \min_\theta \bar{E}\varepsilon^2(t,\theta) = \lambda^2$$

and all solutions satisfy $\hat{\theta} \in D_T$.

Case II: Closed loop experiment. Then u & e are not independent.

You can write

$$\bar{V}(\theta) - \bar{V}(\theta_o) = E((\varepsilon(t,\theta) - \varepsilon(t,\theta_o))\varepsilon(t,\theta_o)) + \frac{1}{2}E\left((\varepsilon(t,\theta) - \varepsilon(t,\theta_o))^2\right)$$

Notice that

$\varepsilon(t,\theta) - \varepsilon(t,\theta_o)$ is indep of $\varepsilon(t,\theta_o) = e_o$

$\bar{E}(\varepsilon(t,\theta) - \varepsilon(t,\theta_o))^2 = E(\Delta E)^2 \quad > 0$

if $\theta \neq \theta_o$ & Z^∞ is informative.

$\Rightarrow \theta = \theta_o$ is the minimizer(s) $(\hat{\theta} \in D_T)$.

Independently Parametrized Set

$$m: \quad y(t) = G(q,p)u + H(q,\eta)e \quad , \quad \theta = \begin{bmatrix} \rho \\ \eta \end{bmatrix}$$

$$D_G = \{\rho | G(q,\rho) = G_o\} \neq \phi$$

No assumptions are made on the noise model

Z^∞ is informative (eq. u is p.e.)

$\widehat{\theta}_N = \begin{bmatrix} \widehat{\rho}_N \\ \widehat{q}_N \end{bmatrix}$ is the estimate.

$$\theta \in D_m$$

Thm: $\widehat{\rho}_N \to D_G.$

Proof: $\varepsilon(t,\theta) = H^{-1}(q,\eta)[y(t) - G(q,\rho)u]$

$$= H^{-1}(q,\eta)[(G_o - G(q,\rho))u + H_o e]$$

$$= u_F(t,\eta,\rho) + e_F(t,\eta)$$

$$\bar{V}(\theta) = \bar{E}\left(\varepsilon^2(t,\theta)\right) = \bar{E}{u_F}^2 + \bar{E}{e_F}^2 \geq \bar{E}{e_F}^2$$

$$\min_{\theta} \bar{V}(\theta) = \min_{e,\eta}\left(\bar{E}{u_F}^2 + E{e_F}^2\right).$$

$$|G_o - G(q,\rho)|^2\Phi_u = 0 \quad \Rightarrow \quad G(q,p) = G_o \quad \Rightarrow \quad \rho \in D_G$$

Frequency Domain Interpretation of the Limit

$$\bar{V}(\theta) = \frac{1}{2}\bar{E}\varepsilon^2 = \frac{1}{4\pi}\int_{-\pi}^{\pi}\Phi_\varepsilon(\omega,\theta)d\omega$$

$$\varepsilon(t,\theta) = H^{-1}(q,\theta)[(G_o - G)u + v_o] \qquad v_o = H_o e$$

$$\Phi_\varepsilon(\omega,\theta) = \frac{\left|G_o - G\left(e^{i\omega},\theta\right)\right|^2 \Phi_u(\omega) + \Phi_{v_o}}{\left|H\left(e^{i\omega},\theta\right)\right|^2}$$

$$\bar{V}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\left|G_o - G\left(e^{i\omega},\theta\right)\right|^2 \Phi_u(\omega) + \Phi_{v_o}}{\left|H\left(e^{i\omega},\theta\right)\right|^2}dw$$

Case I: Suppose $H\left(e^{i\omega},\theta\right) = H^*\left(e^{i\omega}\right) = \text{ fixed}$

$$\text{argmin } \bar{V}(\theta) = \text{argmin } \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\left|G_o - G\left(e^{i\omega},\theta\right)\right|^2 \Phi_u(\omega)}{\left|H^*\left(e^{i\omega},\theta\right)\right|^2} dw$$

Best approx. of G_o, with a weight given by the signal to noise ratio.

Case II: Independently parametrized set

If $\theta^* = \begin{bmatrix} \rho^* \\ \eta^* \end{bmatrix}$ is a minimizer, then

$$\rho^* = \text{argmin } \int_{-\pi}^{\pi} \frac{\left|G_o - G\left(e^{i\omega},\rho\right)\right|^2 \Phi_u(\omega)}{\left|H^*\left(e^{i\omega},\eta^*\right)\right|^2} dw$$

$$\eta^* = \text{argmin} \int_{-\pi}^{\pi} \frac{\Phi_{ER}(\omega, \rho^*)}{\left|H^*\left(e^{i\omega}, \eta\right)\right|^2} dw$$

where $\Phi_{ER}(\omega, \rho^*) = \text{spectrum } y - G(q, \rho^*)u$

$$= \left|G_o\left(e^{i\omega}\right) - G\left(e^{i\omega}, \rho^*\right)\right| \Phi_u + \Phi_v(\omega)$$

(re-write) $= \lambda^* \left|N(\omega, \rho^*)\right|^2$ (spectral fact.)

$$\frac{\Phi_{ER}(\omega, \rho^*)}{\left|H^*\left(e^{i\omega}, \eta\right)\right|^2} = \lambda^* \left|1 + \frac{N(\omega, \rho^*) - H\left(e^{i\omega}, \eta\right)}{H\left(e^{i\omega}, \eta\right)}\right|^2$$

$$= \lambda^* \left[1 + \left|R\left(e^{i\omega}, \eta\right)\right|^2 + R\left(e^{i\omega}, \eta\right) + \bar{R}\left(e^{i\omega}, \eta\right)\right]$$

with $R = \dfrac{N(\omega, \rho^*) - H\left(e^{i\omega}, \eta\right)}{H\left(e^{i\omega}, \eta\right)}$

Since both $\boldsymbol{N}$ & $\boldsymbol{H}$ are monic, H^{-1} is stable. Then,

$$R = \sum_{k>1} r(k)e^{-i\omega k} \quad \& \quad r(0) = \int_{-\pi}^{\pi} R(\omega, \eta)dw \equiv 0$$

$$\Rightarrow \int \frac{\Phi_{ER}(\omega, \rho^*)}{\left|H\left(e^{i\omega}, \eta\right)\right|^2} = \int \lambda^* \left[1 + \left|R\left(e^{i\omega}, \eta\right)\right|^2\right] dw$$

$$\Rightarrow \eta^* = \text{argmin} \int_{-\pi}^{\pi} \lambda^* \left|R\left(e^{i\omega}, \eta\right)\right|^2 dw$$

$$= \text{argmin} \int_{-\pi}^{\pi} \lambda^* \left|\frac{1}{N\left(e^{i\omega}, \rho^*\right)} - \frac{1}{H\left(e^{i\omega}, \eta\right)}\right|^2 \Phi_{ER}(\omega, \rho^*)dw$$

η^* is chosen such that $\frac{1}{H(e^{i\omega},\eta)}$ resembles $\frac{1}{N(e^{i\omega},\rho^*)}$; the inverse of the spectral factor of Φ_{ER}.

Equivalently, $H\left(e^{i\omega}, \eta\right)$ approximates the spectral factor of the error spectrum in the class of admissible $\boldsymbol{H}$'s.