FUNDAMENTAL PROPERTIES

- **Identifiability**
- **Persistency of Excitation**
- **Consistency of PEM**
- **Asymptotic Distribution of PEM**

Reminder: Identifiability

Def (4.1): A *predictor model* of an LTI system is a stable filter $W(q)$ that defines a predictor

$$\hat{y}_{t|t-1} = W(q)\begin{bmatrix} u_t \\ y_t \end{bmatrix}$$

Def (4.2): A *complete probabilistic model* of an LTI system is a pair (W, p_w) of a predictor model W and the probability density function p_w of the prediction error

REMINDER: IDENTIFIABILITY (CONT.)

Def: A *model set* is a collection of models

$$\mathcal{M}^* := \{W_\alpha(q): \quad \alpha \in \mathcal{A}\}, \qquad \mathcal{A}: \text{index set}$$

Def (4.3): A *model structure* $\mathcal{M}$ is a parameterization of a model set $\mathcal{M}^*$, i.e. a smooth mapping from a connected open set $D_{\mathcal{M}} \subset \mathbb{R}^d$ to a model set $\mathcal{M}^*$,

$$\theta \in D_{\mathcal{M}} \mapsto \mathcal{M}(\theta) = W_{\alpha(\theta)}(q) = W(q,\theta) \in \mathcal{M}^*$$

such that the gradients

$$\Psi(z,\theta) := \frac{d\mathcal{M}(\theta)}{d\theta} = \frac{\partial}{\partial\theta} W(q,\theta)$$

are stable

REMINDER: IDENTIFIABILITY (CONT.)

Def (4.6-4.8): A model structure is

- *globally identifiable* at $\theta^* \in D_{\mathcal{M}}$ if

$$W(q,\theta) = W(q,\theta^*), \quad \theta \in D_{\mathcal{M}} \qquad \Rightarrow \qquad \theta = \theta^*$$

- *locally identifiable* at $\theta^* \in D_{\mathcal{M}}$ if for some $\varepsilon > 0$

$$W(q,\theta) = W(q,\theta^*), \quad \theta \in D_{\mathcal{M}}, \quad \|\theta - \theta^*\| < \varepsilon \qquad \Rightarrow \qquad \theta = \theta^*$$

- *globally (locally) identifiable* if it is globally (locally) identifiable for almost all $\theta^* \in D_{\mathcal{M}}$

- *strictly globally (locally) identifiable* if it is globally (locally) identifiable for all $\theta^* \in D_{\mathcal{M}}$

Reminder: Identifiability (cont.)

What about the "true" parameters?

Parameterized structure: $y_t = G(q;\theta)u_t + H(q;\theta)w_t$

True system ($\mathcal{S}$)*:* $y_t = G_0(q)u_t + H_0(q)w_t$

Possibilities: "$\mathcal{S} \in \mathcal{M}$" (i.e. there exists a $\theta_0 \in D_{\mathcal{M}}$ such that $W(q,\theta_0) \in \mathcal{M}^*$)
"$\mathcal{S} \notin \mathcal{M}$" (*undermodelling*)

Def: $D_T(\mathcal{S},\mathcal{M}) := \{\theta \in D_{\mathcal{M}} : \quad G_0(z) = G(z;\theta),\ H_0(z) = H(z;\theta) \text{ a.e. } (z \in \mathbb{C})\}$

If $\mathcal{S} \in \mathcal{M}$ and $\mathcal{M}$ is globally identifiable at $\theta = \theta_0$ then $D_T(\mathcal{S},\mathcal{M}) = \{\theta_0\}$

INFORMATIVE DATA

Does the data set $Z^\infty := \{u_1, y_1, u_2, y_2, \ldots\}$ *allow us to distinguish between models?*

Def (8.1): A quasi-stationary data set Z^∞ is *informative enough* w.r.t. $\mathcal{M}^*$ if

$$\bar{E}\{\left([W_1(q) - W_2(q)]z_t\right)^2\} = 0, \quad W_1, W_2 \in \mathcal{M}^* \quad \Rightarrow \quad W_1(e^{j\omega}) = W_2(e^{j\omega}) \ \text{a.e.}\ (\omega)$$

where $z_t := [u_t^T \quad y_t^T]^T$

Thm (8.1): Z^∞ is informative enough (w.r.t. $\mathcal{L}^*$={LTI models}) if $\Phi_z(e^{j\omega}) > 0$ a.e. ω

Proof:

$$0 = \bar{E}\{[\Delta W(q) z_t]^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Delta W(e^{j\omega})\Phi_z(e^{j\omega})\Delta W^H(e^{j\omega})d\omega \quad \Rightarrow \quad \Delta W(e^{j\omega}) = 0 \ \text{a.e.}(\omega)$$

PERSISTENCE OF EXCITATION

Which conditions on $\{u_t\}$ ensure informative data Z^∞?

Motivating example: SISO open loop case

$$\bar{E}\{([W_1(q)-W_2(q)]z_t)^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\left[\left|\frac{G_1}{H_1}-\frac{G_2}{H_2}\right|^2 + \left|\frac{1}{H_2}-\frac{1}{H_1}\right|^2 |G_0|^2\right]\Phi_u + \lambda^2\left|\frac{1}{H_2}-\frac{1}{H_1}\right|^2 |H_0|^2\right\}$$

By assumption $H_i(e^{j\omega}) \neq 0$ for all ω, so $\bar{E}\{([W_1(q)-W_2(q)]z_t)^2\} = 0$ implies (a.e.(ω))

$$H_1(e^{j\omega}) = H_2(e^{j\omega})$$

$$\left|G_1(e^{j\omega}) - G_2(e^{j\omega})\right|^2 \Phi_u(\omega) = 0 \qquad (*)$$

Needed: Conditions on Φ_u under which $(*) \Rightarrow G_1(e^{j\omega}) = G_2(e^{j\omega})$ a.e.(ω)

$\Rightarrow$ **Persistency of excitation**

PERSISTENCE OF EXCITATION (CONT.)

Def (13.1): A quasi-stationary signal $\{u_t\}$ is *persistently exciting (p.e.) of order* n if for all filters

$$M_n(q) = m_1 q^{-1} + \cdots + m_n q^{-n}$$

we have

$$\left|M_n(e^{j\omega})\right|^2 \Phi_u(\omega) \equiv 0 \qquad \Rightarrow \qquad M_n(e^{j\omega}) \equiv 0$$

Remark: For SISO models, this is equivalent to say that $\Phi_u(\omega) > 0$ for at least n distinct frequencies

Def (13.2): A quasi-stationary signal $\{u_t\}$ is *persistently exciting* if $\Phi_u(\omega) > 0$ a.e. (ω)

PERSISTENCE OF EXCITATION (CONT.)

Lemma (13.1): $\{u_t\}$ is p.e. of order n iff

$$\bar{R}_n := \begin{bmatrix} R_u(0) & \cdots & R_u(n-1) \\ \vdots & \ddots & \vdots \\ R_u(n-1) & \cdots & R_u(0) \end{bmatrix} > 0$$

Proof: Let $v_t := M_n(q)u_t = [m_1 \quad \cdots \quad m_n][u_{t-1} \quad \cdots \quad u_{t-n}]^T$. Then

$$\left|M_n(e^{j\omega})\right|^2 \Phi_u(\omega) \equiv 0 \quad \Leftrightarrow \quad \bar{E}\{v_t v_t^T\} = [m_1 \quad \cdots \quad m_n]\bar{R}_n[m_1 \quad \cdots \quad m_n]^T = 0$$

Also, $\bar{R}_n > 0$ iff

$$[m_1 \quad \cdots \quad m_n]\bar{R}_n[m_1 \quad \cdots \quad m_n]^T = 0 \qquad \Rightarrow \qquad m_1 = \cdots = m_n = 0$$

■

PERSISTENCE OF EXCITATION (CONT.)

Examples:

1. Step input: p.e. of order 1, but not of greater order

2. White noise: p.e. (of every order n)

3. ARMA: p.e. (of every order n)

4. Multisine: p.e. of order equal to number of spectral lines in $(-\pi, \pi]$

PERSISTENCE OF EXCITATION (CONT.)

Relation between p.e. and informative data:

Thm (13.1): Let $\mathcal{M}$ be a model structure of rational LTI models such that

$$G(q,\theta) = \frac{B(q,\theta)}{F(q,\theta)} = \frac{b_1 q^{-nk} + \cdots + b_{nb} q^{-nk-nb+1}}{1 + f_1 q^{-1} + \cdots + f_{nf} q^{-nf}}$$

Then, in open loop, if $\{u_t\}$ is p.e. of order $nb + nf$, Z^{∞} is informative enough w.r.t. $\mathcal{M}$

CONSISTENCY OF PEM

PEM: $\hat{\theta}_N(Z^N) := \arg\min_{\theta \in D_{\mathcal{M}}} V_N(\theta, Z^N), \qquad Z^N := [u_1 \;\; y_1 \; \cdots \; u_N \;\; y_N]$

where typically: $$V_N(\theta, Z^N) := \frac{1}{2N} \sum_{t=1}^{N} \varepsilon_t^2(\theta)$$

Question: Does $\hat{\theta}_N(Z^N)$ converge a.s. (as $N \to \infty$), and to what?

Idea: Study the convergence of $\hat{\theta}_N(Z^N)$ from the uniform convergence of $V_N(\theta, Z^N)$:

$$\sup_{\theta \in D_{\mathcal{M}}} \left| V_N(\theta, Z^N) - \bar{V}(\theta) \right| \xrightarrow[N \to \infty]{a.s.} 0$$

where usually: $$\bar{V}(\theta) = \frac{1}{2} \bar{E} \varepsilon_t^2(\theta)$$

CONSISTENCY OF PEM (CONT.)

System Assumptions:

$$y_t = G_0(q)u_t + H_0(q)w_t$$
$$u_t = -F(q)y_t + r_t$$

where:

- the closed loop is internally stable
- there is at least one time delay in FG_0
- H_0 is monic and inversely stable
- $\{w_t\}$ is a sequence of independent random variables of zero mean, variance λ_0 and bounded moments of order $4+\delta$ (for some $\delta > 0$)
- $\{r_t\}$ is quasi-stationary

CONSISTENCY OF PEM (CONT.)

Model Assumptions:

Def: A family of filters $\{G_\alpha(q):\ \alpha \in \mathcal{A}\}$, with $G_\alpha(q) = \sum_{k=1}^{\infty} g_k^\alpha q^{-k}$, is *uniformly stable* if there is a sequence $\{g_k\}$ such that

$$\begin{cases} \left| g_k^\alpha \right| \le g_k, & \text{for all } \alpha \in \mathcal{A}, \quad k \in \mathbb{N} \\ \displaystyle\sum_{k=1}^{\infty} g_k < \infty \end{cases}$$

Def (8.3): A model structure $\mathcal{M}$ is uniformly stable if $D_{\mathcal{M}}$ is compact and $\{W(q,\theta),\ \Psi(q,\theta),\ (d/d\theta)\Psi(q,\theta):\ \theta \in D_{\mathcal{M}}\}$ is uniformly stable

CONSISTENCY OF PEM (CONT.)

Convergence:

Thm (8.2): Under the system assumptions, and if $\mathcal{M}$ is uniformly stable:

$$\hat{\theta}_N \xrightarrow[N\to\infty]{a.s.} D_c := \arg\min_{\theta\in D_{\mathcal{M}}} \bar{V}(\theta) = \left\{\theta \in D_{\mathcal{M}} : \quad \bar{V}(\theta) = \min_{\tilde{\theta}\in D_{\mathcal{M}}} \bar{V}(\tilde{\theta})\right\}$$

Interpretation:

$\hat{\theta}_N$ will converge to the best *approximation* of the system that is available in $\mathcal{M}^*$

CONSISTENCY OF PEM (CONT.)

Example (8.1): *Bias in ARX models*

System: $y_t + a_0 y_{t-1} = b_0 u_{t-1} + w_t + c_0 w_{t-1}$

$\{u_t\}$ is white

ARX model: $\hat{y}_{t|t-1} = -a y_{t-1} + b u_{t-1}, \qquad \theta = [a \quad b]^T$

Then:

$$\bar{V}(\theta) = \bar{E}\{[y_t + a y_{t-1} - b u_{t-1}]^2\}$$
$$= r_0(1 + a^2 - 2 a a_0) + b^2 - 2 b b_0 + 2 a c_0, \qquad r_0 = E\{y_t^2\}$$

so $\hat{\theta}_N \xrightarrow[N\to\infty]{a.s.} \theta^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix} = \begin{bmatrix} a_0 - \dfrac{c_0}{r_0} \\ b_0 \end{bmatrix} \neq \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}$ ($\hat{\theta}_N$ is *asymptotically biased*)

However: $\bar{V}(\theta^*) = 1 + c_0^2 - \dfrac{c_0^2}{r_0^2} < 1 + c_0^2 = \bar{V}(\theta_0)$

Consistency of PEM (cont.)

Consistency:

If $\mathcal{S} \in \mathcal{M}$, when does $\hat{\theta}_N \to \theta_0$? We need to force $D_c = D_T = \{\theta_0\}$

Thm (8.3): Suppose

- system assumptions
- $\mathcal{M}$ uniformly stable
- $\mathcal{S} \in \mathcal{M}$
- $\mathcal{M}$ globally identifiable at θ_0
- Z^∞ informative enough with respect to $\mathcal{M}^*$

Then: $\hat{\theta}_N \xrightarrow[N\to\infty]{a.s.} \theta_0$

FREQUENCY DOMAIN INTERPRETATION OF $\bar{V}(\theta)$

Main Tool: *Parseval Relation*

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

Now,

$$\varepsilon_t(\theta) = H_\theta^{-1}\begin{bmatrix}G_0 - G_\theta & H_0 - H_\theta\end{bmatrix}\begin{bmatrix}u_t \\ w_t\end{bmatrix} + w_t$$

In open loop:

$$\boxed{\bar{V}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\left|G_0 - G_\theta\right|^2\frac{\Phi_u}{\left|H_\theta\right|^2} + \frac{\lambda_0^2}{\left|H_\theta\right|^2}\left|H_0 - H_\theta\right|^2\right\}d\omega + \frac{\lambda_0}{2}}$$

FREQUENCY DOMAIN INTERPRETATION OF $\bar{V}(\theta)$ (CONT.)

Some interesting cases:

1. For a fixed noise model $H_\theta = H_*$:

$$\bar{V}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\left|G_0 - G_\theta\right|^2 \frac{\Phi_u}{\left|H_*\right|^2}\right\} d\omega + \text{constant}$$

Weighting function: $Q_* = \Phi_u \big/ \left|H_*\right|^2$

For OE models with $\Phi_u(\omega) = 1$ and $H_*(\omega) = 1$, θ^* gives equally good fit for all ω

2. If $H_0(\omega) = 0$ and we use an ARX model with $\Phi_u(\omega) = 1$:

$$\bar{V}(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\left|G_0 - G_\theta\right|^2 \left|A\right|^2\right\} d\omega + \text{constant} \Rightarrow$$ $\left|A\right|^2$ usually penalizes high-frequency misfit much more!

ASYMPTOTIC DISTRIBUTION OF PEM

Thm (9.1): Suppose

- system assumptions
- $\mathcal{M}$ linear and uniformly stable
- $\mathcal{S} \in \mathcal{M}$ and $\hat{\theta}_N$ is consistent
- $\bar{V}''(\theta_0) > 0$

Then,
$$\boxed{\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow[N\to\infty]{d} \mathcal{N}(0, P_{\theta_0})}$$

where
$$\boxed{P_{\theta_0} = \lambda_0[\bar{E}\{\Psi_t(\theta_0)\Psi_t^T(\theta_0)\}]^{-1} = \lambda_0[\bar{V}''(\theta_0)]^{-1}}$$

P_{θ_0} is the C-R bound for Gaussian $\{w_t\}$, so in that case PEM is asymptotically efficient

Remark: To conclude that $\sqrt{N}\operatorname{cov}\hat{\theta}_N \xrightarrow[N\to\infty]{} P_{\theta_0}$ we need stronger assumptions on $\{w_t\}$ and $\bar{V}(\theta)$

ASYMPTOTIC DISTRIBUTION OF PEM (CONT.)

By Parseval,

$$\boxed{P_{\theta_0} = \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\Phi_v} T' \Phi_{\chi_0} T'^H d\omega \right]^{-1}}$$

where

$$T' = \frac{\partial}{\partial \theta} [G \quad H], \qquad \Phi_{\chi_0} = \begin{bmatrix} \Phi_u & \Phi_{uw} \\ \Phi_{wu} & \lambda_0 \end{bmatrix}, \qquad \Phi_v = \lambda_0 \left| H_0 \right|^2$$

Asymptotic Distribution of PEM (cont.)

Asymptotic Variance of G and H: *Delta Method*

For $N \gg 1$: $$f(\hat{\theta}_N) \approx f(\theta_0) + f'(\theta_0)(\hat{\theta}_N - \theta_0)$$

Therefore:

$$\begin{aligned}\operatorname{cov}\{f(\hat{\theta}_N)\} &= E\{[f(\hat{\theta}_N) - f(\theta_0)][f(\hat{\theta}_N) - f(\theta_0)]^T\} \\ &\approx f'(\theta_0)\operatorname{cov}\{\hat{\theta}_N\} f'^T(\theta_0)\end{aligned}$$

Applying this to $f(\theta) = [G_\theta(e^{j\omega}) \quad H_\theta(e^{j\omega})]^T$ gives

$$\boxed{\operatorname{cov}\begin{bmatrix}\hat{G}_N(e^{j\omega}) \\ \hat{H}_N(e^{j\omega})\end{bmatrix} \approx \frac{1}{N} T'^H \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{\Phi_v} T' \Phi_{\chi_0} T'^H d\omega\right]^{-1} T'}$$