8. SYSTEM IDENTIFICATION

Consider

$$\begin{cases} \dot{X}(t) = A(\vartheta)X(t) + B(\vartheta)U(t) \\ X(0) = X_0 , \end{cases} \qquad \text{(SYS)}$$

where $X(t) \in \mathbb{R}^n$, $U(t) \in \mathbb{R}^p$,

$$\{ Y(t) = C(\vartheta)X(t) + D(\vartheta)U(t)$$

where $Y(t) \in \mathbb{R}^q$ and $\vartheta \in \mathbb{R}^m$ is a parameter vector, assumed to be deterministic. The output $Y(t)$ is measured with a sampling period Δ and affected by noise as follows

$$Z(k\Delta) = Y(k\Delta) + G N_k$$

where $\{N_k\} \in \mathcal{N}(0, I)$ and white (i.e. uncorrelated)

The problem is to estimate ϑ from the measurements $Z(0), Z(\Delta), \dots, Z(k\Delta)$.

First of all,

$$X(t) \triangleq X(t, X_0, U, \vartheta)$$
$$= e^{A(\vartheta)t} X_0 + \int_0^t e^{A(\vartheta)(t-s)} B(\vartheta) U(s) ds .$$

If

$$h(k\Delta, X_0, U, \vartheta) \triangleq C(\vartheta) X(k\Delta, X_0, U, \vartheta)$$
$$+ D(\vartheta) U(k\Delta)$$

$\Rightarrow$

$$Z(k\Delta) = h(k\Delta, X_0, U, \vartheta) + G N_k$$

$$k = 0, \dots, N.$$

Define

$$H_N(X_0, U, \vartheta) \triangleq \begin{bmatrix} h(0, X_0, U, \vartheta) \\ h(\Delta, X_0, U, \vartheta) \\ \vdots \\ h(N\Delta, X_0, U, \vartheta) \end{bmatrix}$$

and

$$Z_N \triangleq \begin{pmatrix} z(0) \\ z(\Delta) \\ \vdots \\ z(N\Delta) \end{pmatrix}, \quad E_N \triangleq \begin{bmatrix} GN_0 \\ GN_1 \\ \vdots \\ GN_N \end{bmatrix}$$

$\Rightarrow$

$$\boxed{Z_N = H_N(x_0, U, \vartheta) + E_N}$$

is the <u>measurement equation</u>.

Since ϑ is deterministic, the more convenient way of performing an estimate of ϑ is the maximum likelihood criterium.

First, we will assume x_0 known.

The density p_{Z_N} of Z_N is equal to (recall that E_N is gaussian)

$$p_{Z_N}(z_N, \vartheta) = \frac{1}{(2\pi)^{q(N+1)/2} \det^{1/2} \Psi_N} \cdot e^{-\frac{1}{2}(z_N - \overline{z_N}(\vartheta))^T \Psi_N^{-1} (z_N - \overline{z_N}(\vartheta))}$$

where

$$\overline{z_N}(\vartheta) = E\{z_N\} = H_N(x_0, U, \vartheta)$$

$$\Psi_N(\vartheta) = E\{(z_N - \overline{z_N}(\vartheta))(z_N - \overline{z_N}(\vartheta))^T\}$$

$$= E\left\{\begin{pmatrix} GN_0 \\ \vdots \\ GN_N \end{pmatrix}(N_0^T G^T \dots N_N^T G^T)\right\}$$

$$= \begin{pmatrix} GE\{N_0 N_0^T\}G^T & \dots & GE\{N_0 N_N^T\}G^T \\ \vdots & & \\ GE\{N_N N_0^T\}G^T & \dots & GE\{N_N N_N^T\}G^T \end{pmatrix}$$

$$= \begin{pmatrix} GG^T & 0 & \dots & 0 \\ 0 & GG^T & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & GG^T \end{pmatrix}$$

If we use the logarithm likelihood function

$$J_N(\vartheta) = \ln p_{Z_N}(z_N, \vartheta) \qquad (C)$$

the maximum likelihood estimate of ϑ is given by

$$\hat{\vartheta}_N = \arg\max_{\vartheta} J_N(\vartheta)$$

$$= \arg\min \frac{1}{2}\left(z_N - H_N(X_0, U, \vartheta)\right)^T \Psi_N^{-1} \cdot$$
$$\cdot \left(z_N - H_N(X_0, U, \vartheta)\right)$$

In general, this problem is solved numerically, with more or less important complexity.

If X_0 is not known, it can be modeled as a random vector with mean $\bar{X}_0$ and covariance Ψ_{X_0}.

It is reasonable to assume X_0 to be uncorrelated with GN_k, $\forall k$.

We can rewrite the vector H_N by separating the term depending from X_0 and U:

$$H_N(X_0, U, \vartheta) = S_N(\vartheta) X_0 + F_N(U, \vartheta)$$

where

$$S_N(\vartheta) = \begin{pmatrix} C(\vartheta) \\ C(\vartheta)e^{A(\vartheta)\Delta} \\ \vdots \\ C(\vartheta)e^{A(\vartheta)N\Delta} \end{pmatrix}$$

$$F_N(U,\vartheta) = \begin{bmatrix} D(\vartheta)U(0) \\ \int_0^{\Delta} e^{A(\vartheta)(\Delta-s)} B(\vartheta)U(s)ds + D(\vartheta)u(\Delta) \\ \vdots \\ \int_0^{N\Delta} e^{A(\vartheta)(N\Delta-s)} B(\vartheta)U(s)ds + D(\vartheta)u(N\Delta) \end{bmatrix}$$

The measurement equation is now:

$$Z_N = S_N(\vartheta)X_0 + F_N(U,\vartheta) + E_N$$

Moreover,

$$E\{Z_N\} = S_N(\vartheta)\overline{X}_0 + F_N(U,\vartheta) .$$

If $\tilde{X}_0 \triangleq X_0 - \overline{X}_0$:

$$\Psi_{Z_N}(\vartheta) = E\{(Z_N - E\{Z_N\})(Z_N - E\{Z_N\})^T\}$$

$$= E\{(S_N(\vartheta)\tilde{X}_0 + E_N)(S_N(\vartheta)\tilde{X}_0 + E_N)^T\}$$

$$= S_N(\vartheta)\Psi_{X_0}S_N^T(\vartheta) + \Psi_N$$

(by independence of $\tilde{X}_0$ and E_N)

The logarithmic likelihood function is:

$$J_N(\vartheta) = \ln p_{Z_N}(z_N, \vartheta)$$

$$= \overbrace{\frac{1}{(2\pi)^{q(N+1)/2} \det^{1/2} \Psi_N(\vartheta)}}^{(I)} \cdot$$

$$\cdot \underbrace{e^{-\frac{1}{2}(z_N - S_N(\vartheta)x_0 - F_N(U,\vartheta))^T \Psi_N^{-1}(\vartheta)(z_N - S_N(\vartheta)x_0 - F_N(U,\vartheta))}}_{(II)}$$

Notice that in this case

$$\hat{\vartheta}_N = \arg\max_{\vartheta} J_N(\vartheta)$$

$$\neq \arg\min_{\vartheta} \frac{1}{2}(z_N - S_N(\vartheta)x_0 - F_N(U,\vartheta)) \Psi_N^{-1}(\vartheta) \cdot$$

$$\cdot (z_N - S_N(\vartheta)x_0 - F_N(U,\vartheta))$$

since the term (I) depends on ϑ. Therefore, the problem is computationally more complex. However, it is important, before solving the above optimization problems, to understand if the estimation problem is well-posed. This leads to the notions of parametric identifiability.

8.1. OUTPUT DISTINGUISHABILITY AND IDENTIFIABILITY

Let the output of the system be represented as

$$Y(t) = h(t, X_0, U, \vartheta), \quad \vartheta \in \Theta.$$

By defining EXPERIMENT the pair (X_0, U), where $X_0 \in \mathbb{R}^n$ and $U \in \mathcal{U}_T$, a set of input functions defined over $[0, T]$, let

$$\mathcal{E}_T \triangleq \{(X_0, U) : U \in \mathcal{U}_T\}.$$

DEFINITION. The values $(\bar{\vartheta}, \bar{\alpha})$ of a parameter $\vartheta \in \Theta$ is INDISTINGUISHABLE from the output Y with respect to $\mathcal{E}_T$ if

$$h(t, X_0, U, \bar{\vartheta}) = h(t, X_0, U, \bar{\alpha})$$

$\forall (X_0, U) \in \mathcal{E}_T$, $\forall t \in [0, T]$.

$\mathcal{U}_T$ is assumed to be the set of input functions, uniformly bounded over $[0, T]$. Let $S(\vartheta, \epsilon)$ be the spheric neighbourhood centered at ϑ with radius ϵ.

DEFINITION. A parameter $\vartheta \in \Theta$ is LOCALLY IDENTIFIABLE (around $\bar{\vartheta}$) if there exists $\varepsilon > 0$ such that all pairs $(\bar{\vartheta}, \bar{\alpha})$, for which $\bar{\alpha} \in S(\bar{\vartheta}, \epsilon) \cap \Theta$ and $\bar{\alpha} \neq \bar{\vartheta}$, are distinguishable. The parameter $\vartheta \in \Theta$ is GLOBALLY IDENTIFIABLE if $(\bar{\vartheta}, \bar{\alpha})$ is distinguishable $\forall\, \bar{\alpha} \in \Theta$ with $\bar{\alpha} \neq \bar{\vartheta}$ and all values $\bar{\vartheta}$ of ϑ.

The following result gives a sufficient condition for indistinguishability.

<u>FACT.</u> A pair (ϑ, α), $\vartheta, \alpha \in \Theta$, is indistinguishable with respect to $\mathcal{E}_T$, $T>0$, with $x_0 = 0$, if and only if

$$D(\vartheta) = D(\alpha)$$

$$C(\vartheta) A^{\ell}(\vartheta) B(\vartheta) = C(\alpha) A^{\ell}(\alpha) B(\alpha)$$

$$\ell = 0, 1, 2, \ldots \qquad (D)$$

$(\Rightarrow)$ If (D) holds, since the impulsive output response (with parameter β) is

$$w(t, \beta) = C(\beta) e^{A(\beta)t} B(\beta) + D(\beta) \delta_0(t)$$

and since

$$C(\beta) e^{A(\beta)} B(\beta) = \sum_{i=0}^{\infty} C(\beta) A^{i}(\beta) B(\beta) \frac{t^{i}}{i!},$$

we obtain

$$w(t, \beta) = w(t, \alpha)$$

$$\Rightarrow h(t, 0, u, \vartheta) = h(t, 0, u, \alpha).$$

$(\Leftarrow)$ If (ϑ, α) are indistinguishable with $x_0 = 0$:

$$C(\vartheta)\int_0^t e^{A(\vartheta)(t-\tau)}B(\vartheta)u(\tau)d\tau$$

$$+ D(\vartheta)u(t)$$

$$= C(\alpha)\int_0^t e^{A(\alpha)(t-\tau)}B(\alpha)u(\tau)d\tau$$

$$+ D(\alpha)u(t),$$

$\forall u \in \mathcal{U}_T, \forall t \in [0, T]$.

For $t = 0 \Rightarrow D(\vartheta) = D(\alpha)$

$$\Rightarrow \int_0^t [C(\vartheta)e^{A(\vartheta)(t-\tau)}B(\vartheta) - C(\alpha)e^{A(\alpha)(t-\tau)}B(\alpha)] \cdot$$

$$\cdot u(\tau)d\tau = 0, \quad \forall t \in [0, T].$$

By Euler-Lagrange lemma:

f continuous & $\int_0^t f(\tau)g(\tau)d\tau = 0$

$\forall g$ continuous $\Rightarrow f(\tau) \equiv 0$

It follows

$$C(\vartheta)e^{A(\vartheta)t}B(\vartheta) = C(\alpha)e^{A(\alpha)t}B(\alpha)$$

$\forall t \in [0, T]$

and from this (calculating the derivatives at $t=0$)

$$C(\vartheta)B(\vartheta) = C(\alpha)B(\alpha)$$
$$C(\vartheta)A(\vartheta)B(\vartheta) = C(\alpha)A(\alpha)B(\alpha)$$
$$\vdots$$
$$C(\vartheta)A^{\ell}(\vartheta)B(\vartheta) = C(\alpha)A^{\ell}(\alpha)B(\alpha),$$

$\forall\, \ell = 0, 1, \ldots$

Actually, it is not necessary to check (D) for all $l=0,1,\ldots$ but only for $\ell = 0, 1, \ldots, 2n-1$.

<u>FACT</u>. Let $p=q=1$ (single input, single output) with $(A(\vartheta), B(\vartheta))$ reachable and $(C(\vartheta), A(\vartheta))$ observable. The pair (ϑ, α) is indistinguishable if and only if (D) holds for $\ell = 0, 1, \ldots, 2n-1$.

Proof. We have only to prove that (D) for $\ell = 0, 1, \ldots, 2n-1$ imply (D) for all $\ell \geq 0$. By Cayley–Hamilton theorem

$$\left.\begin{aligned} A^{n+k}(\vartheta) &= \sum_{i=0}^{n-1} a_i A^{i+k}(\vartheta) \\ A^{n+k}(\alpha) &= \sum_{i=0}^{n-1} b_i A^{i+k}(\alpha) \end{aligned}\right\} k = 0, 1, \ldots$$

where a_i and b_i are the coefficients of the characteristic polynomials of $A(\vartheta)$ and $A(\alpha)$.

Premultiplying by $C(\cdot)$ and postmultiplying by $B(\cdot)$

$$C(\vartheta) A^{n+k}(\vartheta) B(\vartheta) = \sum_{i=0}^{n-1} a_i C(\vartheta) A^{i+k}(\vartheta) B(\vartheta)$$
$$C(\alpha) A^{n+k}(\alpha) B(\alpha) \sum_{i=0}^{n-1} b_i C(\alpha) A^{i+k}(\alpha) B(\alpha) \qquad \text{(E)}$$

$k = 0, 1, \ldots, \ldots$

We prove now that if (D) holds for $l = 0, 1, .., 2n-1$ then the characteristic polinomials of $A(\vartheta)$ and $A(\alpha)$ are the same. Consider (E) for $k = 0, .., n-1$ and subtract one from the other, assuming (D) for $l = 0, 1, .., 2n-1$:

$$\sum_{i=0}^{n-1} (a_i - b_i) C(\vartheta) A^{i+k}(\vartheta) B(\vartheta) = 0 \qquad k = 0, .., n-1$$

This equality can be rewritten as

$$\underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}_{\text{nonsingular}} \underbrace{\begin{bmatrix} B & AB & \dots & A^{n-1}B \end{bmatrix}}_{\text{nonsingular}} \begin{bmatrix} a_0 - b_0 \\ a_1 - b_1 \\ \vdots \\ a_{n-1} - b_{n-1} \end{bmatrix} = 0$$

$$\Rightarrow \quad a_i = b_i \quad \forall\, i = 0, .., n-1.$$

But from (E) and since $a_i = b_i$, $i = 0, \ldots, n-1$

it follows

$$C(\vartheta)A^{n+k}(\vartheta)B(\vartheta) =$$

$$C(\alpha)A^{n+k}(\alpha)B(\alpha)$$

$\forall\ k > n-1$

(D) requires to check $2n+1$ equalities.

If $(A(\vartheta), B(\vartheta), C(\vartheta))$ is not reachable/observable, the number of equalities to be checked out is even less.

<u>FACT</u> Let $p = q = 1$, m_ϑ and m_α be the dimensions of the reachable and observable subspace for the representations $(A(\vartheta), B(\vartheta), C(\vartheta))$ and, respectively, $(A(\alpha), B(\alpha), C(\alpha))$. Let $m = \max\{m_\vartheta, m_\alpha\}$. The pair (ϑ, α) is indistinguishable if and

only if

$$\boxed{\begin{aligned} &D(\vartheta) = D(\alpha) \\ &C(\vartheta)A^{\ell}(\vartheta)B(\vartheta) = C(\alpha)A^{\ell}(\alpha)B(\alpha) \\ &\ell = 0, 1, \dots, 2m-1. \end{aligned}}$$

Proof. Consider the Kalman decomposition of $(A(\vartheta), B(\vartheta), C(\vartheta))$

$$\bar{A} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} & \bar{A}_{14} \\ 0 & \bar{A}_{22} & 0 & \bar{A}_{24} \\ 0 & 0 & \bar{A}_{33} & \bar{A}_{34} \\ 0 & 0 & 0 & \bar{A}_{44} \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \\ 0 \\ 0 \end{bmatrix}$$

$$\bar{C} = \begin{bmatrix} 0 & \bar{C}_2 & 0 & \bar{C}_4 \end{bmatrix}$$

with $\bar{A}_{22}$ $(m \times m)$. By construction,

$$C(\vartheta)A^{\ell}(\vartheta)B(\vartheta) = \bar{C}_2(\vartheta)\bar{A}_{22}^{\ell}(\vartheta)\bar{B}_2(\vartheta)$$

$\ell = 0, \dots, n$.

Moreover, m is equal to the rank

of the Hankel matrix:

$$\begin{bmatrix} C(\vartheta)B(\vartheta) & C(\vartheta)A(\vartheta)B(\vartheta) & \cdots \\ C(\vartheta)A(\vartheta)B(\vartheta) & C(\vartheta)A^2(\vartheta)B(\vartheta) & \cdots \\ C(\vartheta)A^2(\vartheta)B(\vartheta) & \cdots & \cdots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

and $m \le n$.

Also for $p \geq 1$ and $q \geq 1$ we have the following result.

<u>FACT</u>. Let $(C(\vartheta), A(\vartheta))$ osservabile and $(A(\vartheta), B_j(\vartheta))$ reachable for at least one j. The pair (ϑ, α) is indistinguishable if and only if

$$D(\vartheta) = D(\alpha)$$

$$C(\vartheta)A^\ell(\vartheta)B(\vartheta) = C(\alpha)A^\ell(\alpha)B(\alpha)$$

$\ell = 0, 1, \ldots, 2n-1$.

Proof. The proof is as in the case $p=q=1$. We obtain

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} [B \ AB \ \ldots \ A^{n-1}B] \begin{bmatrix} (a_0-b_0)I_p \\ (a_1-b_1)I_p \\ \vdots \\ (a_{n-1}-b_{n-1})I_p \end{bmatrix}$$

$= 0$ (I_p is the $p \times p$ identity matrix)

Since $\operatorname{rank}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} = q \Rightarrow$

$$[B \ AB \ \ldots \ A^{n-1}B] \begin{bmatrix} (a_0-b_0)I_p \\ \vdots \\ (a_{n-1}-b_{n-1})I_p \end{bmatrix} = 0$$

If $R_j \triangleq [B_j \ AB_j \ \ldots \ A^{n-1}B_j]$:

$$R_j \begin{bmatrix} a_0-b_0 \\ \vdots \\ a_{n-1} \ b_{n-1} \end{bmatrix} = 0$$

But R_j is nonsingular for at least one j (by assumption) $\Rightarrow$

$a_0 = b_0, \ldots, a_{n-1} = b_{n-1} \Rightarrow$

the characteristic polynomials of $A(\vartheta)$ and $A(\alpha)$ are equal. The proof follows from here as in the case $p=q=1$ ◄

We can state now an important result for local identifiability of the parameter ϑ.

FACT. Let $p=1$ and

$$R(\vartheta) \triangleq \begin{pmatrix} D(\vartheta) \\ C(\vartheta)B(\vartheta) \\ C(\vartheta)A(\vartheta)B(\vartheta) \\ \vdots \\ C(\vartheta)A^{2n-1}(\vartheta)B(\vartheta) \end{pmatrix}.$$

If the components of $R(\vartheta)$ are continuosly differentiable w.r.t. ϑ, $\vartheta \in \Theta \subseteq \mathbb{R}^m$, ϑ is locally identifiable if $\frac{\partial R}{\partial \vartheta}$ has rank $= m$

<u>Proof</u> There exists a neighbourhood $\mathcal{U}$ of ϑ for which $R(\vartheta)$ is bijective and, therefore $R(\vartheta) \neq R(\alpha)$ for all $\alpha \in \mathcal{U} \Rightarrow (\alpha, \vartheta)$ is not indistinguishable $\forall \alpha \in \mathcal{U}$. ◀

In the case $p \geq 1$, we define

$$R_j(\vartheta) \triangleq \begin{pmatrix} D_j(\vartheta) \\ C(\vartheta) B_j(\vartheta) \\ C(\vartheta) A(\vartheta) B_j(\vartheta) \\ \vdots \\ C(\vartheta) A^{2n-1}(\vartheta) B_j(\vartheta) \end{pmatrix}$$

and check that

$$\operatorname{rank} \frac{\partial R_j(\vartheta)}{\partial \vartheta} = m$$

for at least one j. In this case, ϑ is locally identifiable.

8.2. CONSISTENT PARAMETRIC ESTIMATION

Let's get back to the problem of
maximizing the cost function (C)
(pg. 239) in order to obtain a
maximum likelihood estimate $\hat{\vartheta}_N$ of ϑ.
A necessary condition for the consi-
stency of $\hat{\vartheta}_N$ is that ϑ be locally
identifiable. However, local identifia-
bility is not sufficient to guarantee
the consistency of $\hat{\vartheta}_N$. Also the system's
inputs have an important role under
this regard. If the system is at rest
($X(0)=0$) with zero input, the output
is zero and it is clear that any
parameter in this situation cannot
be estimated, even if it is locally
identifiable. Also, in the steady state
regime of the system with constant
input, it is not possible to estimate

dynamic parameters such as the time constants of the system. Under this regard, the notion of persistently exciting input has a very important role in guaranteeing consistent estimates.

For defining persistent excitation, first consider scalar parameters ϑ. Let

$$A_\vartheta \triangleq \frac{dA}{d\vartheta}, \quad B_\vartheta \triangleq \frac{dB}{d\vartheta}$$

$$C_\vartheta \triangleq \frac{dC}{d\vartheta}, \quad D_\vartheta \triangleq \frac{dD}{d\vartheta}$$

$$\left(\frac{\text{SENSITIVITY OF } A, B, C, D}{\text{w.r.t. } \vartheta} \right)$$

and Y_ϑ

$$\left(\frac{\text{SENSITIVITY OF } Y}{\text{w.r.t. } \vartheta} \right)$$

generated by the system:

$$\begin{bmatrix} \dot{X} \\ \dot{X}_\vartheta \end{bmatrix} = \begin{bmatrix} A & 0 \\ A_\vartheta & A \end{bmatrix} \begin{bmatrix} X \\ X_\vartheta \end{bmatrix} + \begin{bmatrix} B \\ B_\vartheta \end{bmatrix} U \qquad (SS)$$

$$Y_\vartheta = \begin{bmatrix} C_\vartheta & C \end{bmatrix} \begin{bmatrix} X \\ X_\vartheta \end{bmatrix} + D_\vartheta U$$

<u>REMARK</u>. The system (SS) is defined by the matrices

$$\bar{A} \triangleq \begin{bmatrix} A & 0 \\ A_\vartheta & A \end{bmatrix}, \quad \bar{B} \triangleq \begin{bmatrix} B \\ B_\vartheta \end{bmatrix}$$

$\bar{C} \triangleq [C_\vartheta \; C]$, $\bar{D} \triangleq D_\vartheta$. Notice:

$$\bar{D} = D_\vartheta = \frac{d}{d\vartheta} D$$

$$\bar{C}\bar{B} = C_\vartheta B + C B_\vartheta = \frac{d}{d\vartheta}(CB)$$

$$\bar{C}\bar{A}\bar{B} = C_\vartheta A B + C A_\vartheta B + C A B_\vartheta = \frac{d}{d\vartheta}(CAB)$$

$$\vdots$$

$$\bar{C}\bar{A}^k\bar{B} = \frac{d}{d\vartheta}(CA^kB) \qquad \forall k \geq 0$$

The condition of local identifiability for ϑ is (with $m=1$: pg. 254)

$$\frac{dR}{d\vartheta} \neq 0 \quad , \quad R(\vartheta) = \begin{bmatrix} D \\ CB \\ CAB \\ \vdots \\ CA^{2n-1}B \end{bmatrix}$$

Since

$$\left.\begin{array}{l} \bar{D} = 0 \\ \bar{C}\bar{B} = 0 \\ \bar{C}\bar{A}\bar{B} = 0 \\ \vdots \\ \bar{C}\bar{A}^{2n-1}\bar{B} = 0 \end{array}\right\} \Leftrightarrow \text{the impulsive output response } W(t) \text{ of (SS) is identically zero}$$

it follows that

$$\begin{cases} \vartheta \text{ is locally identifiable if and only if} \\ W(t) \equiv 0 \end{cases}$$

The output Y will contain information on ϑ if $Y_\vartheta(t) \neq 0$. For this, it is sufficient that $U(t) \neq 0$. However, this is not sufficient for the consistency of the estimate $\hat{\vartheta}_N$ of ϑ.

DEFINITION. An input $U(t)$, uniformly bounded on $[0,+\infty)$, is PERSISTENTLY EXCITING (w.r.t. a parameter $\vartheta \in \mathbb{R}$) if for each $\alpha > 0$ there exist $\delta, \beta > 0$:

$$\int_t^{t+\delta} \| U(\tau) \|^2 d\tau \geq \alpha, \quad \forall t \geq 0$$

$$\int_t^{t+\delta} \| Y_\vartheta(\tau) \|^2 d\tau \geq \beta, \quad \forall t \geq 0$$

P.S. $\| Y_\vartheta(\tau) \|^2$ can be replaced by $Y_\vartheta^T(\tau) W Y_\vartheta(\tau)$ where W is symmetric and positive definite.

(*)

A necessary condition for persistent excitation is the local identifiability of ϑ (because otherwise $Y_\vartheta(\tau) \equiv 0$).

In the case of vector parameter ($m > 1$) it is still possible to define the sensitivity of A, B, C, D and Y as in the case $m = 1$, recalling that if $m > 1$ X_ϑ is a $n \times m$ matrix and Y_ϑ a $q \times m$ matrix.

(*) An equivalent definition is obtained by replacing $\int_t^{t+\delta} \; \forall t \geq 0$ with $\frac{1}{T} \int_0^T \; \forall T \geq T_0$.

The condition of local identifiability is in the case $m > 1$ equivalent to the linear independence of the columns of

$$\begin{bmatrix} \bar{D} \\ \bar{C}\bar{B} \\ \bar{C}\bar{A}\bar{B} \\ \vdots \\ \bar{C}\bar{A}^{2n-1}\bar{B} \end{bmatrix} \qquad \text{(i.e. rank} = m).$$

The notion of persistent excitation is modified as follows for $m > 1$.

<u>DEFINITION</u>. An input $U(t)$, uniformly bounded over $[0, +\infty)$, is PERSISTENTLY EXCITING w.r.t. ϑ if for each symmetric and positive definite $W \in \mathbb{R}^{q \times q}$ there exist $\delta, \beta > 0$:

$$\int_t^{t+\delta} \| u(\tau) \|^2 \geq \alpha, \qquad \forall t \geq 0$$

$$\int_t^{t+\delta} Y_\vartheta^T(\tau) W Y_\vartheta(\tau) d\tau \geq \beta I, \forall t \geq 0.$$

With discrete observations $Y(t_i)$ we say that an input U is persistently exciting if there exist an integer N_0 and $\bar{\alpha}, \bar{\beta} > 0$:

$$\begin{cases} \frac{1}{N} \sum_{i=0}^{N} \| U(t_i) \|^2 \geq \bar{\alpha}, & \forall N > N_0 \\ \frac{1}{N} \sum_{i=1}^{N} Y_\vartheta^T(t_i) W Y_\vartheta(t_i) \geq \bar{\beta} I, & \forall N > N_0. \end{cases}$$

The matrix W is conveniently taken as $(GG^T)^{-1}$ where GG^T is the covariance of the measurement error.

FACT. Let ϑ^* be the true value of ϑ.
Assume:
i) ϑ^* be locally identifiable,
ii) the system (SYS) (pg. 236) be asymptotically stable,
iii) A, B, C, D are continuous w.r.t. ϑ, with their first and second derivatives,
(iv) the inputs U are uniformly bounded and persistently exciting

$\Rightarrow$ the maximum likelihood estimate of ϑ is consistent.

To prove this result, define

$$\gamma(N,\vartheta) \triangleq \left[\frac{d J_N(\vartheta)}{d\vartheta}\right]^T$$

$$M(N,\vartheta) \triangleq \frac{d}{d\vartheta}\gamma(N,\vartheta)$$

$$Q(N,\vartheta) \triangleq \frac{d}{d\vartheta} M(N,\vartheta)$$

The element $m_{i,j}$ of $M(N,\vartheta)$ is

$$m_{i,j} = \left(M(N,\vartheta)\right)_{i,j} = \frac{\partial^2 J_N(\vartheta)}{\partial\vartheta_j \partial\vartheta_i}$$

and the element $q_{i,j,k}$ of $Q(N,\vartheta)$ is

$$q_{i,j,k} = \left[Q(N,\vartheta)\right]_{i,j,k} = \frac{\partial m_{i,j}}{\partial\vartheta_k}$$

$$= \frac{\partial^3 J_N(\vartheta)}{\partial\vartheta_k \partial\vartheta_j \partial\vartheta_i}$$

Given $\Delta\vartheta \in \mathbb{R}^m$ the product $Q(N,\vartheta)\Delta\vartheta$ is a square matrix with entry (i,j):

$$\left(Q(N,\vartheta)\Delta\vartheta\right)_{i,j} = \sum_{k=1}^{m} q_{i,j,k}\,\Delta\vartheta_k$$

We have

$$\gamma(N,\vartheta)-\gamma(N,\vartheta^*)=\int_0^1 \frac{d}{ds}\gamma(N,s\vartheta+(1-s)\vartheta^*)ds$$

$$=\int_0^1 M(N,s\vartheta+(1-s)\vartheta^*)ds\cdot(\vartheta-\vartheta^*)$$

and similarly

$$M(N,s\vartheta+(1-s)\vartheta^*)-M(N,\vartheta^*)$$

$$=\int_0^1 \frac{dM}{ds}(N,\tau[s\vartheta+(1-s)\vartheta^*]+(1-\tau)\vartheta^*)d\tau$$

$$=\int_0^1 e(N,s\tau\vartheta+(1-s\tau)\vartheta^*)d\tau\cdot[s(\vartheta-\vartheta^*)].$$

Defining

$$\bar{e}(N,\vartheta)\triangleq\int_0^1\int_0^1 e(N,s\tau\vartheta+(1-s\tau)\vartheta^*)d\tau\cdot s\,ds$$

and replacing the last equality into the first one:

$$\gamma(N,\vartheta)=\gamma(N,\vartheta^*)+M(N,\vartheta^*)(\vartheta-\vartheta^*)$$
$$+[\bar{e}(N,\vartheta)(\vartheta-\vartheta^*)](\vartheta-\vartheta^*) \quad (T)$$

which is the Taylor expansion of $r(N,\vartheta)$ at $\vartheta = \vartheta^*$, truncated at the first order term.

The maximum likelihood estimate $\hat{\vartheta}_N$ is such that

$$\boxed{r(N,\vartheta) = \left[\frac{dJ_N(\vartheta)}{d\vartheta}\right]^T_{\vartheta=\hat{\vartheta}_N} = 0}$$

But

$$r(N,\vartheta) = -2\left(\frac{d}{d\vartheta}H_N(\vartheta)\right)^T \Psi_N^{-1}\left(Z_N - H_N(\vartheta)\right)$$

$$M(N,\vartheta) = 2\left(\frac{d}{d\vartheta}H_N(\vartheta)\right)^T \Psi_N^{-1}\frac{d}{d\vartheta}H_N(\vartheta) + M_1(N,\vartheta)$$

$\boxed{\#1.}$ Under assumptions (ii), (iii) and uniform boundedness of the inputs U:

$$P-\lim_{N\to+\infty}\left\|\frac{1}{N}M(N,\vartheta^*) - M_1(N,\vartheta^*)\right\| = 0$$

#2. Under the same assumptions of #1 :

$$P-\lim_{N \to +\infty} \frac{1}{N} \gamma(N, \vartheta^*) = 0$$

#3. Under the same assumptions of #1, given $\varepsilon > 0$ there exist $N'(\varepsilon)$ and $c > 0$ such that for all $N > N'(\varepsilon)$ and $\vartheta \in S(\vartheta^*, r)$, the sphere with radius r centered at ϑ^* for which the system is asymptotically stable $\forall \vartheta \in S(\vartheta^*, r)$, we have with probability $\geq 1 - \varepsilon$:

$$\left\| \frac{1}{N} \left[\bar{Q}(N, \vartheta)(\vartheta - \vartheta^*) \right] (\vartheta - \vartheta^*) \right\| \leq c \, \| \vartheta - \vartheta^* \|^2 .$$

Using persistent excitation of U (ii) on page 262), we find N_0 and $\sigma > 0$ such that

$$\frac{1}{N} M_1(N, \vartheta^*) \geq \sigma I$$

$$\forall N > N_0$$

It follows from #1:

#4. Given $\varepsilon', \varepsilon$ sufficiently small there exists $N''(\varepsilon, \varepsilon')$ such that with probability $\geq 1-\varepsilon$:

$$\frac{1}{N} M(N, \vartheta^*) \geq (\sigma - \varepsilon) \cdot I$$

$$\forall N > N''$$

By choosing $\varepsilon' = \frac{\sigma}{2}$, in probability with $N > N''$

$$\left\| \left[\frac{1}{N} M(N, \vartheta^*)\right]^{-1} \right\| \leq \frac{1}{\sigma'}, \quad \sigma' \triangleq \frac{\sigma}{2}$$

To complete the proof, define

$$\nu \triangleq \vartheta - \vartheta^*$$

and

$$T_N(\nu) \triangleq \nu - M^{-1}(N, \vartheta^*)\,\gamma(N, \vartheta^* + \nu)$$

Using (T) on pg. 264 in $T_N(\nu)$:

$$T_N(\nu) = \nu - M^{-1}(N, \vartheta^*) \cdot$$

$$\cdot \left(\gamma(N, \vartheta^*) + M(N, \vartheta^*)\nu + [\bar{\varrho}(N, \vartheta^* + \nu)\nu]\nu\right)$$

$$= -M^{-1}(N, \vartheta^*)\,\gamma(N, \vartheta^*)$$

$$- M^{-1}(N, \vartheta^*)[\bar{\varrho}(N, \vartheta^* + \nu)\nu]\nu$$

$$= -\left[\frac{M(N, \vartheta^*)}{N}\right]^{-1}\left(\frac{\gamma(N, \vartheta^*)}{N} + \left[\frac{1}{N}\bar{\varrho}(N, \vartheta^* + \nu)\nu\right]\nu\right)$$

Using #2, #3 and #4, given $\varepsilon > 0$, $\delta_1 > 0$ (with δ_1 : $\frac{2\delta_1}{\sigma'} \leq \frac{\sigma'}{4c}$), we find $\bar{N}$:

$\forall\, N > \overline{N}$ with probability $\geq 1-\varepsilon$:

$$\| \Gamma_N(\nu) \| \leq \frac{\delta_1}{\sigma'} + \frac{c}{\sigma'} \|\nu\|^2 .$$

Let $\delta \in \left[\frac{2\delta_1}{\sigma'}, \frac{\sigma'}{4c} \right] \Rightarrow$

$$\begin{cases} \frac{2\delta_1}{\sigma'} \leq \delta \Rightarrow \frac{\delta_1}{\sigma'} \leq \frac{\delta}{2} \\ \delta \leq \frac{\sigma'}{4c} \Rightarrow \frac{4c\delta}{\sigma'} \leq 1 \Rightarrow \frac{c\delta^2}{\sigma'} \leq \frac{\delta}{4} \end{cases}$$

$\Rightarrow \forall\, \nu : \|\nu\| \leq \delta$ we have

$$\| \Gamma_N(\nu) \| \leq \frac{\delta}{2} + \frac{\delta}{4} = \frac{3\delta}{4}$$

$\Rightarrow \exists\, \hat{\nu}_N : \quad \| \hat{\nu}_N \| \leq \delta$

$$\Gamma_N(\hat{\nu}_N) = \hat{\nu}_N$$

(fixed point theorem).

Since ε and δ_1 are arbitrary,

$\Rightarrow \gamma(N, \vartheta^*) = 0$. This proves that $\hat{\vartheta}_N \to \vartheta^*$ in probability $\blacktriangleleft$