NOTES ON SYSTEM IDENTIFICATION, PREDICTION AND FILTERING

STEFANO BATTILOTTI

MASTER DEGREE IN
CONTROL ENGINEERING
(SAPIENZA UNIVERSITY OF ROME)

RECALLS OF PROBABILITY THEORY

Let

- Ω be the space of elementary events (results of experiments) and $\omega \in \Omega$ be its points
- a σ-algebra $\mathcal{F}$ of events S (subsets of Ω). A σ-algebra satisfies the following property

(i) $\Omega \in \mathcal{F}$ (sure event)

(ii) $S \in \mathcal{F} \Rightarrow S^c \in \mathcal{F}$ (complementary event)

(iii) $S_i \in \mathcal{F}$, $i = 1, 2, \ldots,$

$\Rightarrow \bigcup_i S_i \in \mathcal{F}$

(A2)

Clearly,

- $\Omega^c = \phi \in \mathcal{F}$
- $\bigcap_i S_i = \left(\bigcup_i S_i^c\right)^c \in \mathcal{F}$

Let also on $(\Omega, \mathcal{F})$ be defined a probability measure $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ as follows:

(i) $\mathbb{P}(S) \geq 0 \quad \forall\, S \in \mathcal{F}$

(ii) $\mathbb{P}(\Omega) = 1$

(iii) $\mathbb{P}(\bigcup_i S_i) = \sum_i \mathbb{P}(S_i)$ if $\{S_i\}$ is a sequence of $S_i \in \mathcal{F}$ such that $S_i \cap S_j = \{\phi\}$ for $i \neq j$.

A3

Clearly,

$$1 = \mathcal{P}(\Omega) = \mathcal{P}(S \cup S^c)$$

$$= \mathcal{P}(S) + \mathcal{P}(S^c) \quad \text{so that}$$

$$\boxed{\mathcal{P}(S^c) = 1 - \mathcal{P}(S)}$$

Also, if $S = \Omega \Rightarrow \boxed{\mathcal{P}(\emptyset) = 0}$.

Moreover, if $S_1 \subset S_2$:

$$\mathcal{P}(S_2) = \mathcal{P}((S_2 \setminus S_1) \cup S_1)$$

$$= \mathcal{P}(S_2 \setminus S_1) + \mathcal{P}(S_1) \geq \mathcal{P}(S_1)$$

Therefore

$$\boxed{S_1 \subseteq S_2 \Rightarrow \mathcal{P}(S_1) \leq \mathcal{P}(S_2)}$$

and

$$0 \leq \boxed{\mathcal{P}(S) \leq \mathcal{P}(\Omega) = 1}$$

A4

A RANDOM VECTOR on $(\Omega, \mathcal{F})$ is defined as the function $X : \Omega \to \mathbb{R}^n$ such that

$$\boxed{\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall\, x \in \mathbb{R}^n}$$

Here by $f(\omega) \leq \alpha$ we mean shortly that $f_1(\omega) \leq \alpha_1, \dots, f_n(\omega) \leq \alpha_n$. We say also that X is $\mathcal{F}$-measurable.

To each random vector we can associate the following DISTRIBUTION FUNCTION $F_X : \mathbb{R}^n \to [0,1]$

$$\boxed{F_X(x) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \leq x\}\right)}$$

A5

A standard σ-algebra on $\mathbb{R}^n$ is the so-called $\mathcal{B}^n$, Borel σ-algebra over $\mathbb{R}^n$. The function $F'_X : \mathcal{B}^n \to [0,1]$ (DISTRIBUTION OF X)

$$\boxed{F'_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}), \quad B \in \mathcal{B}^n}$$

is a probability measure over $(\mathbb{R}^n, \mathcal{B}^n)$.

The distribution function F_X satisfies the following important properties:

(i) F_A is a monotone non-decreasing function with respect to each argument x_i, $i=1,..,n$, keeping constant the remaining ones;

(ii) $\lim\limits_{x_i \to -\infty} F_X(x) = 0$, $i=1,..,n$;

(A6)

(iii) $\lim_{x_i \to \infty} F_X(x) = F_{X^{(i)}}(x^{(i)})$

$i = 1, .., n,$

where $X^{(i)} = \begin{pmatrix} X_1 \\ \vdots \\ X_{i-1} \\ X_{i+1} \\ \vdots \\ X_n \end{pmatrix}, x^{(i)} = \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ x_{i+1} \\ \vdots \\ x_n \end{pmatrix};$

(iv) $\lim_{x_1, .., x_n \to \infty} F_X(x) = 1$

(v) $1 - F_X(x) = \mathcal{P}(\{\omega \in \Omega : X(\omega) > x\})$

$\forall\, x \in \mathbb{R}^n$

To each random vector we can also associate a PROBABILITY DENSITY FUNCTION $p_X : \mathbb{R}^n \rightarrow [0, \infty)$

$$\boxed{\mathcal{P}(\{\omega \in \Omega : X(\omega) \leq x\}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p_X(x)\, dx_1 \ldots dx_n}$$

(whenever this equality holds for all $x \in \mathbb{R}^n$).

Clearly,

$$\boxed{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_X(x)\, dx_1 \ldots dx_n = 1}$$

The set $\Omega_A \subset \mathbb{R}^n$ for which

$$\Omega_A = \{ x \in \mathbb{R}^n ;\ p_X(x) > 0 \}$$

is called the SUPPORT of p_X.

If F_X is differentiable then

$$p_X(x) = \left(\frac{\partial^n F_X}{\partial x_1 \ldots \partial x_n} \right)(x)$$

Moreover, if

$$Y = f(X)$$

and X, Y are two random vectors over $(\Omega, \mathcal{F})$ and if $f : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then

$$\boxed{p_Y(y) = p_X(f^{-1}(Y)) \cdot \left| \det \left\{ \frac{\partial f^{-1}}{\partial y} \Big|_{y=Y} \right\} \right|} \tag{A.1}$$

For example

$$Y = MX + m \quad , \quad \det M \neq 0$$

then

$$p_Y(y) = \frac{p_X(M^{-1}(Y - m))}{|\det M|}$$

The EXPECTED VALUE (mean) of a random vector X is defined as follows

$$\boxed{\overline{X} = E[X] = \int_{\Omega_X} x\, p_X(x)\, dx}$$

More generally

$$E[f(X)] = \int_{\Omega_X} f(x)\, p_X(x)\, dx$$

The mean operator satisfies the following properties

- $f(X) = C \Rightarrow E[f(X)] = C$
- $f(X) \gtreqless 0 \Rightarrow E[f(X)] \gtreqless 0$
- $f_1(X) \gtreqless f_2(X) \Rightarrow E[f_1(X)] \gtreqless E[f_2(X)]$

A10

The COVARIANCE MATRIX of X is defined as

$$\boxed{\Psi_X = E\left[(X-\bar{X})(X-\bar{X})^T\right]}$$

The matrix Ψ_X is symmetric and positive semidefinite. Moreover,

$$Tr\,\Psi_X = \sum_{j=1}^{n} E\left[(X_i-\bar{X}_i)^2\right]$$

$$= E\left[\|X-\bar{X}\|^2\right]$$

i.e. the mean quadratic error of X with respect to $\bar{X}$.

For each X_i, $i=1,..,n$,

$\sigma_{X_i}^2 = E\left[(X_i-\bar{X}_i)^2\right]$ is the VARIANCE of X_i while σ_{X_i} is known as standard deviation.

Some peculiar distribution and density functions :

(1) $\boxed{X = c \quad \text{deterministic, known}}$:

$$\begin{cases} F_X(x) = \begin{cases} 0 & x < c \\ 1 & x \geq c \end{cases} \\ \\ p_X(x) = \delta_0(x - c) \end{cases}$$

(δ_0 is the Dirac impulse)

(2) $\boxed{\text{uniform distribution over } \Omega_X}$:

$$\begin{cases} F_X(x) = \dfrac{\text{measure}(\Omega_X \cap (-\infty, x))}{\text{measure}(\Omega_X)} \\ \\ p_X(x) = \begin{cases} \dfrac{1}{\text{measure}\{\Omega_X\}} & x \in \Omega_X \\ 0 & x \notin \Omega_X \end{cases} \end{cases}$$

A12

(iii) $\boxed{\text{gaussian distribution}}$:

$$\begin{cases} F_X(x) = \displaystyle\int_{-\infty}^{x} \frac{1}{(2\pi)^{1/2}} \frac{1}{(\det \Phi)^{1/2}} e^{-\frac{1}{2}(x-c)^T \Phi^{-1}(x-c)} dx \\ p_X(x) = \dfrac{1}{(2\pi)^{1/2}} \cdot \dfrac{1}{(\det \Phi)^{1/2}} e^{-\frac{1}{2}(x-c)^T \Phi^{-1}(x-c)} \end{cases}$$

Note that both density and distribution are uniquely determined by the parameters $\Phi \in \mathbb{R}^{n \times n}$ (symmetric and positive definite) and $c \in \mathbb{R}^n$.

Moreover, it can be seen that

$$\begin{cases} E[X] = c \\ \Psi_X = \Phi \end{cases}$$

If $c = 0$ and $\Phi = I_{n \times n}$ we simply write $X \in N(0, I)$ ("standard gaussian")

A13

If $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \begin{array}{l} \}\, r \\ \}\, s \end{array}$

where X_1, X_2 are random vectors, with density $p_X(x_1, x_2)$ (joint density of X_1 and X_2), then the so-called marginal densities $p_{X_1}(x_1)$ and $p_{X_2}(x_2)$ of X_1 and, respectively, of X_2 are related to $p_X(x_1, x_2)$ as follows

$$\begin{cases} p_{X_1}(x_1) = \displaystyle\int_{\Omega_{X_1}} p_X(x_1, x_2)\, dx_2 \\ p_{X_2}(x_2) = \displaystyle\int_{\Omega_{X_2}} p_X(x_1, x_2)\, dx_1 \end{cases} \quad (A.2)$$

where Ω_{X_1} and Ω_{X_2} and Ω_X are the supports of p_{X_1}, p_{X_2} and respectively p_X and

$$\boxed{\Omega_X = \Omega_{X_1} \times \Omega_{X_2}}$$

Moreover,

$$E[f_1(X_1)] = \int_{\Omega_{X_1}} f_1(x_1)\, p_{X_1}(x_1)\, dx_1$$

$$= \int_{\Omega_{X_1}} f_1(x_1) \left(\int_{\Omega_{X_2}} p_X(x_1, x_2)\, dx_2 \right) dx_1$$

$$= \int_{\Omega_{X_1}} \int_{\Omega_{X_2}} f_1(x_1)\, p_X(x_1, x_2)\, dx_1\, dx_2$$

$$= \int_{\Omega_X} f_1(x_1)\, p_X(x_1, x_2)\, dx = E[f_1(X_1)]$$

In other words, to evaluate the mean of a function only of X_1 it is sufficient the density of X_1.

Therefore,

$$\bar{X} = E[X] = E\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$= \begin{pmatrix} E[X_1] \\ E[X_2] \end{pmatrix} \overset{\text{use } p_{X_1}}{\underset{\text{use } p_{X_2}}{=}} \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix}$$

Moreover,

$$\Psi_X = E\left[(X-\bar{X})(X-\bar{X})^T\right]$$

$$= \left(\begin{array}{c|c} E[(X_1-\bar{X}_1)(X-X_1)^T] & E[(X_1-\bar{X}_1)(X_2-\bar{X}_2)^T] \\ \hline E[(X_2-\bar{X}_2)(X_1-\bar{X}_1)^T] & E[(X_2-\bar{X}_2)(X_2-\bar{X}_2)^T] \end{array}\right)$$

$$= \left(\begin{array}{c|c} \Psi_{X_1} & \Psi_{X_1,X_2} \\ \hline \Psi_{X_2,X_1} & \Psi_{X_2} \end{array}\right)$$

and Ψ_{X_i,X_j} are the MIXED COVARIANCES of X_1 and X_2. If $\Psi_{X_i,X_j} = 0$ then X_i and X_j are said to be UNCORRELATED.

A16

Notice that

$$\Psi_{x_1,x_2} = E[X_1 X_2^T] - \bar{X}_1 \bar{X}_2^T$$

$$\Psi_{x_j} = E[X_j X_j^T] - \bar{X}_j \bar{X}_j^T$$

If X is gaussian also X_1 and X_2 are gaussian with means $\bar{X}_1$, $\bar{X}_2$ and covariances Ψ_{x_1}, Ψ_{x_2}.

The converse is not true (in general, the marginal densities do not determine the joint density).

A17

Given two random vectors X, Y;
We define PROBABILITY DENSITY of X CONDITIONED TO Y the families of densities

$$\boxed{p_{X|Y}(x,y) = \frac{p_{\binom{X}{Y}}(x,y)}{p_Y(y)} \qquad \forall\, y \in \Omega_Y} \qquad (A.3)$$

Here y is a possible value of Y and parametrizes the family of densities for X. It should be noted that

$$\Omega_{X|Y} \subset \Omega_X$$

Actually each $p_{X|Y}(x,y)$, $y \in \Omega_Y$, is a density for X:

$$\int_{\Omega_{X|Y}} p_{X|Y}(x,y)\,dx = \int_{\Omega_X} p_{X|Y}(x,y)\,dx$$

A18

$$= \frac{1}{p_Y(y)} \int_{\Omega_X} p_{\binom{X}{Y}}(x,y)\,dx$$

$$= \frac{p_Y(y)}{p_Y(y)} = 1$$

N.B $p_Y(y) = \int_{\Omega_X} p_{\binom{X}{Y}}(x,y)\,dx$ by definition of marginal density

One can define with the family $p_{X|Y}(x,y)$ the expected value of $f(X)$ conditioned to Y:

$$E[f(X)|Y] = \int_{\Omega_{X|Y}} f(x)\, p_{X|Y}(x,y)\,dx$$

the covariance of X conditioned to Y

$$\Gamma_{X|Y} = E\left[(X - E[X|Y])(X - E[X|Y])^T\right]$$

We list some basic properties of the conditional expectation:

$$(i)\quad E[f_1(X_1) f_2^T(X_2) \mid X_2] = E[f_1(X_1) \mid X_2] f_2^T(X_2) \qquad (A.4)$$

$$(ii)\quad E[E[f(X_1, X_2)] \mid X_2] = E[f(X_1, X_2)] \qquad (A.5)$$

(iii) if X_1, X_2 are gaussian

$$\begin{cases} E[X_1 \mid X_2] = \overline{X}_1 + \Psi_{X_1, X_2} \Psi_{X_2}^{-1} (X_2 - \overline{X}_2) \\ \Psi_{X_1 \mid X_2} = \Psi_{X_1} - \Psi_{X_1, X_2} \Psi_{X_2}^{-1} \Psi_{X_2, X_1} \end{cases}$$

(we will prove this in 2. pg. 65)

A20

Two random vectors X, Y

are INDEPENDENT if

$$\boxed{p_{\binom{X}{Y}}(x,y) = p_X(x)\, p_Y(y)}$$

or equivalently

$$\boxed{F_{\binom{X}{Y}}(x,y) = F_X(x)\, F_Y(y)}$$

If X, Y are independent $\Rightarrow$

$$E[X \mid Y] = E[X]$$

Moreover, if X_1, X_2 are independent $\Rightarrow$

$$E[f_1(X_1)\, f_2(X_2)] = E[f(X_1)]\,E[f(X_2)]$$

$$\boxed{\begin{array}{l}\text{Finally, if } X_1, X_2 \text{ are uncorrelated} \\ \text{and } X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \text{ gaussian} \Rightarrow X_1, X_2 \\ \text{are independent}\end{array}}$$

This follows from (iii) pg. A19 using $\Psi_{X_1 X_2} = \Psi_{X_2 X_1} = 0$. We obtain

$$E[X_1 | X_2] = \bar{X}_1 = E[X_1]$$

$$\Psi_{X_1 | X_2} = \Psi_{X_1}$$

which implies (since $p_{X|Y}$ is gaussian being p_Y and $p\begin{pmatrix} X \\ Y \end{pmatrix}$ gaussian)

$$p_{X_1 | X_2} = p_{X_1}$$

and thus by definition of $p_{X_1|X_2}$

$$p_{\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}}(x_1, x_2) = p_{X_1|X_2}(x_1, x_2)\, p_{X_2}(x_2)$$

$$= p_{X_1}(x_1)\, p_{X_2}(x_2)$$

A sequence of random vectors

$$\{X(j)\},\ j = 1, \dots,$$

is called a RANDOM sequence.

Each $X(j)$ has a density $p_{X(j)}(x_j)$.

If each $p_{X(j)}$ does not depend on j, we say that the sequence is I.I.D. (identically distributed).

If each $x(j)$ is uncorrelated from each $x(i)$, $i \neq j$, we say that the sequence is WHITE. If each $x(j)$ is independent from $x(i)$, $j \neq i$, we say that the sequence is INDIPENDENT.

For a sequence $\{X(j)\}$

$$E[X(i)X^T(j)] = \mathcal{X}_X(i,j)$$

is the AUTOCORRELATION of $\{X(j)\}$.

For two sequence $\{X(j)\}$ and $\{Y(j)\}$

$$E[X(i)Y^T(j)] = \mathcal{X}_{XY}(i,j)$$

is the CROSS-CORRELATION of $\{X(j)\}$ and $\{Y(j)\}$. If $\{X(j)\}$ and $\{Y(j)\}$ are i.i.d. $\mathcal{X}_X$ and $\mathcal{X}_{XY}$ depend only on $i-j$ and not on i and j separately.

Some general notions of CONVERGENCE of sequences $\{X(j)\}$.

(i) $\{X(j)\}$ converges in MEAN p to X if

$$\lim_{j\to\infty} E\{\|X(j)-X\|^p\}=0$$

When $p=2$, we say in QUADRATIC MEAN.

(ii) $\{X(j)\}$ converges in PROBABILITY to X if

$$\lim_{j\to\infty} \mathcal{P}\{\omega\in\Omega : \|X(j)-X\|>\epsilon\}=0 \quad \forall \epsilon>0$$

(iii) $\{X(j)\}$ converges in DISTRIBUTION to X if

$$\lim_{j\to\infty} F_{X(j)} = F_X$$

CONVERGENCE IN P-MEAN

$\Rightarrow$ CONVERGENCE IN PROBABILITY

$\Rightarrow$ CONVERGENCE IN DISTRIBUTION

Finally, the (WEAK) LAW OF LARGE NUMBERS:

Given $\{X(j)\}$ independent, with $\Psi_{X(j)}$ bounded in norm for $j=1,2,\dots$, the random vector

$$\boxed{X_m = \frac{1}{m}\sum_{j=1}^{m} X(j)}$$

is such that

$$\boxed{\lim_{m\to\infty} E\left[\| X_m - \overline{X}_m \|^2\right] = 0} \quad (A.6)$$

If $\{X(j)\}$ is also i.i.d., then

$$\overline{X}_m = E[X(j)]$$

, i.e. the common expected value of each $X(j)$.

ESTIMATION PROBLEM

- θ (parameter vector) $:= \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_\mu \end{pmatrix}$

  unknown (deterministic or random)

- available data:

$$\underbrace{z(j)}_{\text{measurement}} = g(\theta, j) + \underbrace{v(j)}_{\text{noise}} \quad \left(\begin{array}{c}\text{MEASUREMENT} \\ \text{EQUATION}\end{array}\right)$$

$$\boxed{\text{time } j = 1, .., k}$$

- PROBLEM: find an "estimate" or "approximation" $\tilde{\theta}|_k$ of θ as

$$\tilde{\theta}|_k = \gamma(z(1), .., z(k))$$

where γ: [illegible] function

- The estimate $\tilde{\Theta}|_k$ may be constrained to an "admissibility set" $D_\Theta \subset \mathbb{R}^\mu$. Estimates $\tilde{\Theta}|_k \notin D_\Theta$ should be discarded

- available information is given on the noise sequence $\{v(j)\}_{j=1,..,k}$ (A PRIORI INFORMATION) $\longrightarrow$ probability densities $\{p_{v(j)}(n_j)\}_{j=1,..,k}$ and statistical properties of the sequence (whiteness, i.i.d., ...)

- additional information available on Θ if it is random $\longrightarrow$ probability density $p_\Theta(\theta)$. In this case the support $\mathcal{D}_\Theta$ of p_Θ has the role of the admissibility set in the case of deterministic Θ.

Notation:

$$z^{(k)} = \begin{pmatrix} z(1) \\ \vdots \\ z(k) \end{pmatrix}, \quad \nu^{(k)} = \begin{pmatrix} \nu(1) \\ \vdots \\ \nu(k) \end{pmatrix}$$

$$g^{(k)}(\theta) = \begin{pmatrix} g(\theta, 1) \\ \vdots \\ g(\theta, k) \end{pmatrix}$$

The density of $z^{(k)}$ may be derived as follows:

$$p_{z^{(k)}}(\zeta^{(k)}, \theta) = p_{\nu^{(k)}}(\zeta^{(k)} - g^{(k)}(\theta))$$

it depends on θ

or (if θ is random)

$$p_{z^{(k)}|\theta}(\zeta^{(k)}, \theta) = p_{\nu^{(k)}}(\zeta^{(k)} - g^{(k)}(\theta))$$

(see (A.1)).

The importance of $p_Z^{(k)}$ (or $p_{z^{(k)}|\theta}$) relies on its characterization of the "well-posedness" of the estimation problem. The chance of estimating θ is related to the "sensibility" of $p_{z^{(k)}}$ (or $p_{z^{(k)}|\theta}$) with respect to the variations of θ

[RANDOM θ]

DEF. 1 Two vectors $\theta', \theta'' \in \Omega_\theta$ (the support of p_θ), $\theta' \neq \theta''$ are said to be INDISTINGUISHABLE UNDER k MEASUREMENTS if

$$p_{z^{(k)}|\theta}(s^{(k)}, \theta') = p_{z^{(k)}|\theta}(s^{(k)}, \theta'')$$

$\forall\, s^{(k)} \in \Omega_{z^{(k)}|\theta}$ (the support of $p_{z^{(k)}|\theta}$)

[DETERMINISTIC θ]

5

DEF 2. Two vectors $\theta', \theta'' \in D_\theta$ (admissibility set), $\theta' \neq \theta''$, are said INDISTINGUISHABLE UNDER k measurements if

$$p_{z^{(k)}}(z^{(k)}, \theta') = p_{z^{(k)}}(z^{(k)}, \theta'')$$

$\forall\, z^{(k)} \in \Omega_{z^{(k)}}$.

[RANDOM θ]

DEF 1.1 θ IS IDENTIFIABLE UNDER k measurements if there are no indistinguishable pairs (θ', θ''), $\theta' \neq \theta''$, $\theta', \theta'' \in \Omega_\theta$.

[DETERMINISTIC θ]

DEF 2.1 θ is IDENTIFIABLE UNDER k measurements if there are no indistinguishable pairs (θ, θ'), $\theta' \neq \theta$, $\theta' \in D_\theta$.

The property of identifiability of θ is guaranteed by the invertibility of the function $g^{(k)}(\theta)$ in the cases in which:

$$p_{z^{(k)}|\theta}(s^{(k)},\theta) = p_{v^{(k)}}(s^{(k)} - g^{(k)}(\theta))$$

$$\left(\text{or } p_{z^{(k)}}(s^{(k)};\theta) = p_{v^{(k)}}(s^{(k)} - g^{(k)}(\theta))\right)$$

Condition for "local" invertibility of $g^{(k)}(\theta)$ are:

$$\boxed{\operatorname{rank}\left\{ \frac{\partial g^{(k)}}{\partial \theta} \Bigg|_{\theta = \bar\theta} \right\} = p}, \quad \bar\theta \in D_\theta \quad (\text{or } \bar\theta \in \Omega_\theta)$$

(by the implicit function theorem).

A necessary condition for this is

$$\boxed{kq \geq p}$$

If $g^{(k)}(\theta) = C^{(k)}\theta$ (linearity), a necessary and sufficient condition for invertibility of $g^{(k)}(\theta)$ is

$$\boxed{\operatorname{rank} C^{(k)} = p}$$

TIME VARYING PARAMETERS

$\theta(i)$, $i = 1, \dots, k$

If the parameter to be estimated is varying with time, all the above definitions and technical conditions are directly extended by replacing θ with $\theta^{(k)}$, defined below:

NOTATION

$$\theta^{(k)} = \begin{pmatrix} \theta(1) \\ \vdots \\ \theta(k) \end{pmatrix}, \quad g^{(k)}(\theta^{(k)}) = \begin{pmatrix} g(\theta(1), 1) \\ \vdots \\ g(\theta(k), k) \end{pmatrix}$$

In this case, there is the possibility of estimating the parameter $\Theta(i|k)$ at time i using a different number k of measurements.

In particular

- if $i=k \Rightarrow$ FILTERING
- if $i>k \Rightarrow$ PREDICTION
- if $i<k \Rightarrow$ INTERPOLATION

In these case the estimate will be denoted by $\tilde{\Theta}(i|k)$

ESTIMATES PROPERTIES

Some properties have to be introduced in such a way to establish how far an estimate is "good". We consider only the case of constant Θ (random or deterministic) and we leave to the student the extension to time varying $\Theta(i)$.

9. CENTERING

DEF. 3

An estimate $\tilde{\Theta}|_k$ of Θ is said to be CENTERED if

$$\begin{cases} E[\tilde{\Theta}|_k] = \Theta & \text{if } \Theta \text{ is deterministic} \\ E[\tilde{\Theta}|_k] = E[\Theta] & \text{if } \Theta \text{ is random} \\ (\text{or alternatively } E[\tilde{\Theta}|_k \mid \Theta] = \Theta) & \end{cases}$$

The difference

$$E[\tilde{\theta}|_k] - \theta$$
$$(\text{or } E[\tilde{\theta}|_k] - E[\theta])$$

is called POLARIZATION of the estimate.

b. EFFICIENCY

DEF. 4

An estimate $\tilde{\theta}|_k$ of θ is said to be efficient if $\tilde{\theta}|_k$ is centered

and

$$\Psi_{\tilde{\theta}|_k} \leq \Psi_{\hat{\theta}|_k}$$

for all centered estimates $\hat{\theta}|_k$

Note that for a centered $\tilde{\theta}|_k$ if $\tilde{\theta}|_k = \theta - \tilde{\theta}|_k$

$$\Psi_{\tilde{\theta}|_k} = E[(\tilde{\theta}|_k - E[\tilde{\theta}|_k])(\tilde{\theta}|_k - E[\tilde{\theta}|_k])^T] =$$
$$= E[(\tilde{\theta}|_k - \theta)(\tilde{\theta}|_k - \theta)^T] = \Psi_{\tilde{\theta}|_k}$$

In the case of random θ, $\tilde{e}|_k = E[\theta] - \tilde{\theta}|_k$.

c. CONSISTENCY

DEF. 5

$\boxed{\text{An estimate } \tilde{\theta}|_k \text{ of } \theta \text{ is said to be CONSISTENT if } \forall \epsilon > 0 \quad \lim_{k\to\infty} Pr\{\|\tilde{e}|_k\| > \epsilon\} = 0}$

(where $\tilde{e}|_k$ is defined as above).

A sufficient condition for consistency is

$$\boxed{\lim_{k\to\infty} \Psi_{\tilde{e}|_k} = 0}$$

Next, we want to establish conditions for efficiency.

Consider Θ deterministic and assume

$\boxed{H1}$. $p_{Z^{(k)}}(s^{(k)}, \Theta)$ differentiable with respect to Θ near its true value and for each $s^{(k)} \in \mathbb{R}^{kq}$ (here we use the obvious notation $s^{(k)} = (s_1, \dots, s_k)^T$, $s_i \in \mathbb{R}^q$)

$\boxed{H2}$. $$\left\| \frac{\partial p_{Z^{(k)}}(s^{(k)}, \Theta)}{\partial \Theta} \right\| \leq \varphi(s^{(k)}, \Theta)$$

for some φ integrable with respect to $s^{(k)}$ over $\mathbb{R}^{kq}$ uniformly with respect to Θ, i.e.

$$\sup_{\Theta \in D_\Theta} \int_{\mathbb{R}^{kq}} \varphi(s^{(k)}, \Theta)\, ds^{(k)} < \infty$$

Define the FISCHER matrix $R_k(\theta) \in \mathbb{R}^{\mu \times \mu}$ as follows

$$\boxed{R_k(\theta) := E\left[\left(\frac{\partial}{\partial\theta} p_{2^{(k)}}(2^{(k)}, \theta) \right)^T \left(\frac{\partial}{\partial\theta} p_{2^{(k)}}(2^{(k)}, \theta) \right) \right]}$$

If $p_{2^{(k)}}(2^{(k)}, \theta)$ is differentiable two times with respect to θ, we have also

$$\boxed{R_k(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \ln p_{2^{(k)}}(2^{(k)}, \theta) \right]}$$

CRAMER-RAO LOWER BOUND

Assume H1 and H2 and let
$R_k(\theta)$ be nonsingular $\forall \theta \in D_\theta$.
For any centered estimate we have

$$\boxed{\Psi_{\tilde{e}|_k} \geq R_k^{-1}(\theta)} \quad (CR)$$

Proof. First, using H1 and H2

$$E\left\{\theta \frac{\partial \ln p_{z^{(k)}}(z^{(k)},\theta)}{\partial \theta}\right\} = \int_{\Omega_{z^{(k)}}} \theta \frac{\partial p_{z^{(k)}}(s^{(k)},\theta)}{\partial \theta} ds^{(k)}$$

$$= \theta \frac{\partial}{\partial \theta} \int_{\Omega_{z^{(k)}}} p_{z^{(k)}}(s^{(k)},\theta)\, ds^{(k)} = 0 \quad (1)$$

Second, using H1 and H2 and $\tilde{\theta}|_k = \gamma(z^{(k)})$

$$E\left\{\tilde{\theta}|_k \frac{\partial}{\partial \theta} \ln p_{z^{(k)}}(z^{(k)},\theta)\right\} =$$

$$\int_{\Omega_{z^{(k)}}} \gamma(s^{(k)}) \frac{\partial}{\partial \theta} p_{z^{(k)}}(s^{(k)},\theta)\, ds^{(k)}$$

$$= \frac{\partial}{\partial \theta} \int_{\Omega_{z^{(k)}}} \gamma(\varsigma^{(k)})\, p_{z^{(k)}}(\varsigma^{(k)}, \theta)\, d\varsigma^{(k)}$$

$$= \frac{\partial}{\partial \theta} E\left[\gamma(z^{(k)})\right] = \frac{\partial}{\partial \theta} E\left[\tilde{\theta}|_k\right]$$

$$(2)$$

$$= \frac{\partial \theta}{\partial \theta} = I_{\mu \times \mu}$$

(use CENTERING!)

Subtracting (2) from (1)

$$E\left[\tilde{e}|_k \frac{\partial \ln p_{z^{(k)}}(z^{(k)}, \theta)}{\partial \theta}\right] = -I_{\mu \times \mu}$$

Finally consider the matrix

$$E\left\{\left[\tilde{e}|_k + R_k^{-1}(\theta)\left(\frac{\partial}{\partial \theta} \ln p_{z^{(k)}}(z^{(k)}, \theta)\right)\right] \cdot \left[\tilde{e}|_k + R_k^{-1}(\theta)\left(\frac{\partial}{\partial \theta} \ln p_{z^{(k)}}(z^{(k)}, \theta)\right)\right]^T\right\} \quad (3)$$

which is positive semidefinite (≥ 0)

(N.B. Any matrix $G = NN^T \in \mathbb{R}^{n \times n}$ where $N \in \mathbb{R}^n$ is ≥ 0)

Then, taking the products in (3) (16)

$$\Psi_{\tilde{e}|k} + R_k^{-1}(\theta) E\left[\left(\frac{\partial}{\partial\theta}\ln p_{z(k)}(z^{(k)},\theta)\right)^T \tilde{e}_{|k}^T\right]$$

$$+ E\left[\tilde{e}_{|k} \cdot \frac{\partial}{\partial\theta}\ln p_{z(k)}(z^{(k)},\theta)\right] R_k^{-1}(\theta)$$

$$+ R_k^{-1}(\theta) E\left[\left(\frac{\partial}{\partial\theta}\ln p_{z(k)}(z^{(k)},\theta)\right)^T\left(\frac{\partial}{\partial\theta}\ln p_{z(k)}(z^{(k)},\theta)\right)\right] R_k^{-1}(\theta)$$

$$= \Psi_{\tilde{e}|k} - R_k^{-1}(\theta) - R_k^{-1}(\theta) + R_k^{-1}(\theta) R_k(\theta) R_k^{-1}(\theta)$$

$$= \Psi_{\tilde{e}|k} - R_k^{-1}(\theta) \quad \text{which is also} \geq 0$$

It follows (CR). ◁

Recalling that densities transform as follows

$$B = f(A), \quad f \text{ invertible}$$

$$\Downarrow$$

$$p_B(\beta) = p_A(f^{-1}(\beta)) \left| \det\left\{\frac{\partial f^{-1}}{\partial\beta}(\beta)\right\}\right|$$

(see (A.1))

if the measurement equation is given by

$$z^{(k)} = g^{(k)}(\theta) + v^{(k)}$$

(remember $z^{(k)} = (z(1), \dots z(k))^T$
$g^{(k)}(\theta) = (g(\theta, 1), \dots, g(\theta, k))^T$
$v^{(k)} = (v(1), \dots, v(k))^T$)

then

$$\boxed{p_{z^{(k)}}(\zeta^{(k)}, \theta) = p_{v^{(k)}}(\zeta^{(k)} - g^{(k)}(\theta))}$$

(N.B. If θ is random
$p_{z^{(k)}|\theta}(\zeta^{(k)}, \theta) = p_{v^{(k)}}(\zeta^{(k)} - g^{(k)}(\theta))$)

If in addition $\{v(j)\}$ is white and gaussian then $\{v(j)\}$ is also independent $\Longrightarrow$

$$p_{v^{(k)}}(n^{(k)}) = \prod_{j=1}^{k} p_{v(j)}(n_j)$$

and using gaussianity of each $\nu(j)$

$$p_{\nu(j)}(n_j) = \frac{1}{(2\pi)^{\frac{q}{2}} |\det \Psi_{\nu(j)}|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(n_j - E[\nu(j)])^T \Psi_{\nu(j)}^{-1} (n_j - E[\nu(j)])}$$

Therefore, assuming also $E[\nu(j)] = 0$ $\forall j$,

$$\ln p_{z^{(k)}}(\zeta^{(k)}, \theta) = \sum_{j=1}^{k} \frac{1}{(2\pi)^{\frac{q}{2}} |\det \Psi_{\nu(j)}|^{1/2}} - \frac{1}{2} \sum_{j=1}^{k} (\zeta_j - g(\theta, j))^T \Psi_{\nu(j)}^{-1} (\zeta_j - g(\theta, j))$$

Moreover, it can be also seen that also $\{z(j)\}$ is gaussian and white with

$$\boxed{\begin{aligned} E[z^{(k)}] &= g^{(k)}(\theta) \text{ and} \\ \Psi_{z^{(k)}} &= \Psi_{\nu^{(k)}} \end{aligned}}$$

Indeed,

$$p_{a^{(k)}}(S^{(k)},\theta) = p_{\nu^{(k)}}\left(S^{(k)} - g^{(k)}(\theta)\right)$$

$$= \prod_{j=1}^{k} p_{\nu(j)}\left(S_j - g(\theta,j)\right) =$$

$$= \prod_{j=1}^{k} \frac{1}{(2\pi)^{\frac{q}{2}} |\det \Psi_{\nu(j)}|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(S_j - g(\theta,j))^T \Psi_{\nu(j)}^{-1} (S_j - g(\theta,j))}$$

$$= \frac{1}{(2\pi)^{\frac{qk}{2}} |\det \Psi_{\nu^{(k)}}|^{1/2}} \cdot \prod_{j=1}^{k} e^{-\frac{1}{2}(S_j - g(\theta,j))^T \Psi_{\nu(j)}^{-1} (S_j - g(\theta,j))}$$

N.B.
($\det \Psi_{\nu^{(k)}} = \prod_{j=1}^{k} \Psi_{\nu(j)}$ by whiteness and $|\det M_1 \det M_2| = |\det M_1| |\det M_2| = |\det M_1 M_2|$)

$$= \frac{1}{(2\pi)^{\frac{qk}{2}} |\det \Psi_{\nu^{(k)}}|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(S^{(k)} - g^{(k)}(\theta))^T \Psi_{\nu^{(k)}}^{-1} (S^{(k)} - g^{(k)}(\theta))}$$

N.B.
($\Psi_{\nu^{(k)}} = \mathrm{diag}\{\Psi_{\nu(j)}\}$ by whiteness)

which proves that

$z^{(k)}$ is gaussian with

$E[z^{(k)}] = g^{(k)}(\theta)$ and $\Psi_{z^{(k)}} = \Psi_{\nu^{(k)}}$.

Finally, note that under the assumptions that $\{\nu(j)\}$ is gaussian with zero mean (uncorrelation is not needed)

$$\frac{\partial}{\partial\theta} \ln p_{z^{(k)}}(S^{(k)}, \theta) = \left(z^{(k)} - g^{(k)}(\theta)\right)^T \Psi_{\nu^{(k)}}^{-1} \left(\frac{\partial g^{(k)}(\theta)}{\partial\theta}\right);$$

Therefore,

$$R_k(\theta) = E\left[\left(\frac{\partial}{\partial\theta} \ln p_{z^{(k)}}(z^{(k)}, \theta)\right)^T \left(\frac{\partial}{\partial\theta} \ln p_{z^{(k)}}(z^{(k)}, \theta)\right)\right]$$

$$= E\left[\left(\frac{\partial g^{(k)}(\theta)}{\partial\theta}\right)^T \Psi_{\nu^{(k)}}^{-1} \left(z^{(k)} - g^{(k)}(\theta)\right)\left(z^{(k)} - g^{(k)}(\theta)\right)^T \Psi_{\nu^{(k)}}^{-1} \left(\frac{\partial g^{(k)}(\theta)}{\partial\theta}\right)\right]$$

$$= \left(\frac{\partial g^{(k)}(\theta)}{\partial\theta}\right)^T \Psi_{\nu^{(k)}}^{-1} E\left[\left(z^{(k)} - g^{(k)}(\theta)\right)\left(z^{(k)} - g^{(k)}(\theta)\right)^T\right] \Psi_{\nu^{(k)}}^{-1} \left(\frac{\partial g^{(k)}(\theta)}{\partial\theta}\right)$$

$$= \boxed{\left(\frac{\partial g^{(k)}(\theta)}{\partial\theta}\right) \Psi_{\nu^{(k)}}^{-1} \left(\frac{\partial g^{(k)}}{\partial\theta}\right)} \quad (4)$$

If $g^{(k)}(\theta) = C^{(k)}\underline{\theta}$ (LINEARITY) 21

$$\Rightarrow \boxed{R_k(\theta) = C^{(k)^T} \Psi_{v(k)}^{-1} C^{(k)}} \quad (5)$$

For the invertibility of $R_k(\theta)$ in (4) we need the sufficient condition

$$\boxed{\text{rank}\left\{\frac{\partial g^{(k)}}{\partial \theta}(\theta)\right\} = \mu}$$

$\forall \theta$ near its true value

which guarantees local (around the true value of θ) invertibility of $R_k(\theta)$.

In the case of (5) invertibility of $R_k(\theta)$ $\forall\theta$ is guaranteed by

$$\boxed{\text{rank}\, C^{(k)} = \mu}$$

ESTIMATION BASED ON THE FISHER MATRIX $R_k(\theta)$

Let θ be deterministic and $D_\theta = \mathbb{R}^\mu$.

If $R_k(\theta)$ is nonsingular and if for

$$\gamma(\zeta^{(k)}, \theta) := \theta + R_k^{-1}(\theta)\left[\frac{\partial \ln p_{z^{(k)}}(\zeta^{(k)}, \theta)}{\partial \theta}\right]^T$$

we have $\dfrac{\partial \gamma}{\partial \theta}(\zeta^{(k)}, \theta) = 0 \quad \forall\ \zeta^{(k)}, \theta$ then

$$\tilde{\theta}|_k = \theta + R_k^{-1}(\theta)\left[\frac{\partial}{\partial \theta} \ln p_{z^{(k)}}(z^{(k)}, \theta)\right]^T \quad (6)$$

is an centered and efficient estimate of θ and it is the only one with these properties. $\triangleleft$

Proof. (6) is a well-defined estimate of θ since $\dfrac{\partial \gamma}{\partial \theta} \equiv 0$ (γ is independent of θ).

Centering property.

$$E\left[\tilde{\theta}|_k\right] = \theta + R_k^{-1}(\theta) E\left\{\left[\frac{\partial}{\partial\theta} \ln p_{z^{(k)}}(z^{(k)},\theta)\right]^T\right\}$$

$$= \theta + R_k^{-1}(\theta) \int_{\Omega_{z^{(k)}}} \frac{\partial p_{z^{(k)}}(s^{(k)},\theta)}{\partial\theta} \, ds^{(k)}$$

$$= \theta + R_k^{-1}(\theta) \frac{\partial}{\partial\theta} \underbrace{\int_{\Omega_{z^{(k)}}} p_{z^{(k)}}(s^{(k)},\theta) \, ds^{(k)}}_{=1} = \theta$$

Efficiency

$$\Psi_{\tilde{e}|_k} = E\left[(\theta - \tilde{\theta}|_k)(\theta - \tilde{\theta}|_k)^T\right]$$

$$= R_k^{-1}(\theta) E\left\{\left[\frac{\partial}{\partial\theta} \ln p_{z^{(k)}}(z^{(k)},\theta)\right]^T \left[\frac{\partial}{\partial\theta} \ln p_{z^{(k)}}(z^{(k)},\theta)\right]\right\} \cdot$$

$$\cdot R_k^{-1}(\theta) = R_k^{-1}(\theta) R_k(\theta) R_k^{-1}(\theta) = R_k^{-1}(\theta)$$

which is the Cramer-Rao lower bound.

Uniqueness

Suppose $\tilde{\theta}|_k$ be any other centered and efficient estimate of θ. It must be

$$E\{(\theta-\tilde{\theta}|_k)(\theta-\tilde{\theta}|_k)^T\}=E\{\tilde{e}|_k\tilde{e}|_k^T\}$$

$$=R_k^{-1}(\theta) \quad \text{where } \tilde{e}|_k := \theta-\tilde{\theta}|_k .$$

Therefore,

$$0 = E\{\tilde{e}|_k\tilde{e}|_k^T - R_k^{-1}(\theta)\} =$$

$$=E\left\{\left[\tilde{e}|_k+R_k^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\right)^T\right]\cdot\left[\tilde{e}|_k+R_k^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\right)^T\right]^T\right\} \qquad (6)$$

we used here the fact that

$$E\{\tilde{e}|_k \frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\}=-I$$

Taking the trace of both parts in the equality (6)

$$0 = \mathrm{Tr}\, E\left\{\left[\tilde{e}|_k + R_k^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\right)^T\right] \cdot \left[\tilde{e}|_k + R_k^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\right)\right]^T\right\}$$

$$= E\left\{\mathrm{Tr}\left(\left[\tilde{e}|_k + R_k^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\right)^T\right]^T \cdot \left[\tilde{e}|_k + R_k^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\right)^T\right]\right)\right\}$$

we used $\mathrm{Tr}(A \cdot B) = \mathrm{Tr}(B \cdot A)$

$$= E\left\{\left\|\tilde{e}|_k + R_k^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\right)^T\right\|^2\right\}$$

This implies

$$\tilde{e}|_k + R_k^{-1}(\theta)\left(\frac{\partial}{\partial\theta}\ln p_{z^{(k)}}(z^{(k)},\theta)\right) = 0$$

from which uniqueness of $\tilde{\theta}|_k$ $\triangleleft$

CLASSES OF ESTIMATORS

1. WEIGHTED LEAST SQUARES ESTIMATORS

2. MAXIMUM LIKELIHOOD ESTIMATORS

3. BAYESIAN ESTIMATORS

WEIGHTED LEAST SQUARES ESTIMATORS

Let

$$z(j) = C(j)\theta + v(j) \qquad j = 1, \dots, k$$

be the given measurement equation. Assume $E[v(j)] = 0$ $\forall j$. In compact form

$$z^{(k)} = C^{(k)}\theta + v^{(k)}$$

where $C^{(k)} = \begin{pmatrix} C(1) \\ \vdots \\ C(k) \end{pmatrix}$ and $z^{(k)}$ and $v^{(k)}$ as usual.

Define a COST FUNCTION as follows: (θ is deterministic):

$$\boxed{J(\theta) = \left(z^{(k)} - C^{(k)}\theta\right)^T Q^{[k]} \left(z^{(k)} - C^{(k)}\theta\right)}$$

where $Q^{[k]} = \mathrm{diag}\{Q(1), \dots, Q(k)\}$

$$= \begin{pmatrix} Q(1) & & 0 \\ & \ddots & \\ 0 & & Q(k) \end{pmatrix}$$

and $Q(j)$ are symmetric and positive definite matrices, $Q(j) \in \mathbb{R}^{q \times q}$.

Therefore, $Q^{[k]} \in \mathbb{R}^{kq \times kq}$. The matrices $Q(j)$ represents "weights".

Our estimation problem for θ is defined as follows

$$\boxed{\tilde{\theta}\big|_k = \underset{\theta \in \mathbb{R}^{\mu}}{\operatorname{argmin}}\ J(\theta)}$$

N.B.

Since θ must be identifiable we must have $\boxed{\operatorname{rank} C^{(k)} = \mu}$ (7)

To solve this ~~optimization~~ minimization problem we first look for candidates:

$$\left.\frac{\partial J}{\partial \theta}\right|_{\theta = \tilde{\theta}|_k} = 0$$

This gives

$$\left.\frac{\partial J}{\partial \theta}\right|_{\theta=\tilde{\theta}|_k} = -2\left(z^{(k)} - C^{(k)}\tilde{\theta}|_k\right)^T Q^{[k]} C^{(k)}$$
$$= 0$$

which, since $C^{(k)^T} Q^{[k]} C^{(k)}$ is non singular by (7) and $Q^{[k]}$ being invertible, has only one solution

$$\tilde{\theta}|_k = \left(C^{(k)^T} Q^{[k]} C^{(k)}\right)^{-1} C^{(k)^T} Q^{[k]} z^{(k)} \tag{8}$$

By considering

$$\frac{\partial^2 J}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\frac{\partial J}{\partial \theta}\right)^T = 2 I_{\mu \times \mu} > 0$$

It follows that (8)
is indeed the minimum we are seeking. The estimate (8) of θ is called WEIGHTED LEAST SQUARES estimate (WLS).

This estimate is always centered:

$$E[\tilde{\theta}|_k] = (C^{(k)^T} Q^{[k]} C^{(k)})^{-1} C^{(k)^T} Q^{[k]} E[z^{(k)}]$$
$$= (C^{(k)^T} Q^{[k]} C^{(k)})^{-1} C^{(k)^T} Q^{[k]} C^{(k)} \theta$$
$$= \theta$$
$$\uparrow$$
$$\left(\text{since } E[v^{(k)}] = 0 \right)$$

Moreover, the covariance of $\tilde{\theta}|_k$ is evaluated as follows

$$\Psi_{\tilde{\theta}|_k} = \Psi_{\tilde{e}|_k} = E[\tilde{e}|_k \tilde{e}|_k^T]$$

$E[v^{(k)}$

31

$$= \left(C^{(k)^T} Q^{[k]} C^{(k)} \right)^{-1} C^{(k)^T} Q^{[k]} E\left[\nu^{(k)} \nu^{(k)^T} \right] \cdot$$

$$\cdot\, Q^{[k]} C^{(k)} \left(C^{(k)^T} Q^{[k]} C^{(k)} \right)^{-1} =$$

$$= \left(C^{(k)^T} Q^{[k]} C^{(k)} \right)^{-1} C^{(k)^T} Q^{[k]} \Psi_{\nu(k)} Q^{[k]} C^{(k)} \cdot$$

$$\cdot \left(C^{(k)^T} Q^{[k]} C^{(k)} \right)^{-1}$$

If $Q(j) = p\, I_{q \times q}$, $p > 0$, $\forall j$

then we recover the CLASSICAL LEAST SQUARES estimators (i.e. the weights are all equal)

$$\tilde{\theta}\big|_k = \underbrace{\left(C^{(k)^T} C^{(k)} \right)^{-1} C^{(k)^T}}_{= \left(C^{(k)} \right)^{\#} \text{ (pseudoinverse of } C^{(k)})} z^{(k)}$$

with

$$\Psi_{\tilde{\theta}|_k} = \left(C^{(k)^T} C^{(k)} \right)^{-1} C^{(k)^T} \Psi_{\nu(k)} C^{(k)} \left(C^{(k)^T} C^{(k)} \right)^{-1}$$

$$= \left(C^{(k)} \right)^{\#} \Psi_{\nu(k)} \left(\left(C^{(k)} \right)^{\#} \right)^T$$

If the noise sequence $\{\nu(j)\}$ is white (uncorrelated) and each marginal covariance $\Psi_{\nu(j)}$ is nonsingular, we can choose $Q(j) = \Psi_{\nu(j)}^{-1}$. The estimate we obtain in this case is called a MARKOV estimate.

Note that uncorrelation of $\{\nu(j)\}$ allows to set the weights $Q^{[k]} = \Psi_{\nu(k)}^{-1}$ since $Q^{[k]}$ is block diagonal and $\Psi_{\nu(k)} = \mathrm{diag}\{\Psi_{\nu(1)}, \ldots, \Psi_{\nu(k)}\}$ by uncorrelation of $\{\nu(j)\}$. The Markov estimate is then given by

$$\tilde{\theta}|_k = \left(C^{(k)T} \Psi_{\nu(k)}^{-1} C^{(k)}\right)^{-1} C^{(k)T} \Psi_{\nu(k)}^{-1} z^{(k)} \qquad (9)$$

with

$$\Psi_{\tilde{\Theta}|k} = \Psi_{\tilde{e}|k} = \left(C^{(k)T} \Psi_{v|k}^{-1} C^{(k)}\right)^{-1} \quad (10)$$

We want to prove that (9) is the linear and centered estimate which has the ~~minimum~~ covariance among all the linear and centered estimates.

Let
$$\hat{\Theta}|_k = \Lambda^{(k)} z^{(k)}$$
be any linear and centered estimate of Θ. We have the centering property if and only if

$$E\left[\hat{\Theta}|_k\right] = \Lambda^{(k)} E\left[z^{(k)}\right] = \Lambda^{(k)} C^{(k)} \Theta = \Theta$$

or equivalently

$$\Lambda^{(k)} C^{(k)} = I \quad (11)$$

Moreover, if $\hat{e}|_k = \theta - \hat{\theta}|_k$

$$\Psi_{\hat{\theta}|_k} = \Psi_{\hat{e}|_k} = E\left[(\theta - \hat{\theta}|_k)(\theta - \hat{\theta}|_k)^T\right]$$

$$= E\left\{\left(\theta - \Lambda^{(k)}C^{(k)}\theta - \Lambda^{(k)}\nu^{(k)}\right)\left(\theta - \Lambda^{(k)}C^{(k)}\theta - \Lambda^{(k)}\nu^{(k)}\right)^T\right\}$$

$$= \Lambda^{(k)}E\left[\nu^{(k)}\nu^{(k)T}\right]\Lambda^{(k)T} = \Lambda^{(k)}\Psi_{\nu(k)}\Lambda^{(k)T}$$

Check now that $\Psi_{\hat{\theta}|_k} \geq \Psi_{\tilde{\theta}|_k}$

where $\tilde{\theta}|_k$ is given by (9): ~~Using (10)~~

$$\Psi_{\hat{\theta}|_k} - \Psi_{\tilde{\theta}|_k} = \Lambda^{(k)}\Psi_{\nu(k)}\Lambda^{(k)T} - (C^{(k)}\Psi^{-1}_{\nu(k)}C^{(k)})^{-1} =$$

$$= \Lambda^{(k)}\Psi^{1/2}_{\nu(k)}\Psi^{1/2}_{\nu(k)}\Lambda^{(k)T} - \Lambda^{(k)}\Psi^{1/2}_{\nu(k)}\Psi^{-1/2}_{\nu(k)}C^{(k)} \cdot$$

$$\cdot (C^{(k)T}\Psi^{-1}_{\nu(k)}C^{(k)})^{-1} - (C^{(k)T}\Psi^{-1}_{\nu(k)}C^{(k)})^{-1} \cdot$$

$$\cdot C^{(k)T}\Psi^{-1/2}_{\nu(k)}\Psi^{1/2}_{\nu(k)}\Lambda^{(k)T} + (C^{(k)T}\Psi^{-1}_{\nu(k)}C^{(k)})^{-1} \cdot$$

$$\cdot C^{(k)T}\Psi^{-1/2}_{\nu(k)}\Psi^{1/2}_{\nu(k)}C^{(k)}(C^{(k)T}\Psi^{-1}_{\nu(k)}C^{(k)})^{-1}$$

---

we used (11) with the decompositions

$$\Psi^{-1/2}_{\nu(k)}\Psi^{\frac{1}{2}}_{\nu(k)} = I \quad \text{and} \quad \Psi^{1/2}_{\nu(k)}\Psi^{1/2}_{\nu(k)} = \Psi_{\nu(k)}$$

$$= \Lambda^{(k)} \Psi_{\nu(k)}^{1/2} \left\{ \Psi_{\nu(k)}^{1/2} \Lambda^{(k)^T} - \Psi_{\nu(k)}^{-1/2} C^{(k)} \cdot \left(C^{(k)^T} \Psi_{\nu(k)}^{-1} C^{(k)} \right)^{-1} \right\} - \left(C^{(k)^T} \Psi_{\nu(k)}^{-1} C^{(k)} \right)^{-1} \cdot$$

$$\cdot C^{(k)^T} \Psi_{\nu(k)}^{-1/2} \left\{ \Psi_{\nu(k)}^{1/2} \Lambda^{(k)^T} - \Psi_{\nu(k)}^{-1/2} C^{(k)} \cdot \left(C^{(k)^T} \Psi_{\nu(k)}^{-1} C^{(k)} \right)^{-1} \right\} =$$

$$= \left\{ \Lambda^{(k)} \Psi_{\nu(k)}^{1/2} - \left(C^{(k)^T} \Psi_{\nu(k)}^{-1} C^{(k)} \right)^{-1} C^{(k)^T} \Psi_{\nu(k)}^{-1/2} \right\} \cdot \left\{ \underbrace{\Psi_{\nu(k)}^{1/2} \Lambda^{(k)^T} - \Psi_{\nu(k)}^{-1/2} C^{(k)} \left(C^{(k)^T} \Psi_{\nu(k)}^{-1} C^{(k)} \right)^{-1}}_{:= H^{(k)}} \right\}$$

$$= H^{(k)^T} \cdot H^{(k)} \geq 0$$

Therefore $\Psi_{\hat{\Theta}|k} - \Psi_{\tilde{\Theta}|k} \geq 0$

which proves that (9) has the lowest covariance inside the family of $\tilde{\Theta}|_k$. $\triangleleft$

Moreover, (9) is the only estimates with covariance equal to (10). Indeed, if

$$\Psi_{\hat{\theta}|_k} = \Psi_{\tilde{\theta}|_k} \text{ for some}$$

other $\tilde{\theta}|_k$, it should be $H^{(k)} = 0$

or, equivalently,

$$\Lambda^{(k)} = \left(C^{(k)}\Psi^{-1}_{\nu(k)}C^{(k)}\right)^{-1} C^{(k)T}\Psi^{-1}_{\nu(k)}$$

which gives $\hat{\theta}|_k \equiv \tilde{\theta}|_k$.

If in addition, $\{\nu(j)\}$ is GAUSSIAN the estimate $\tilde{\theta}|_k$ given in (9) is also <u>efficient</u>. Indeed, in this case the covariance (10) of (9) is equal to $F_k^{-1}(\theta)$ by the formula (5).

DYNAMIC WEIGHTED LEAST SQUARES ESTIMATION

Here we consider the case of time-varying parameter $\theta(i)$.

The parameter equation is:

(12) $$\theta(j+1) = A(j)\theta(j) + m(j), \quad j=0,1,\dots,k-1$$

where $\theta(j) \in \mathbb{R}^{M}$, $\{m(j)\}$ is a deterministic known sequence, $\theta(0)$ is unknown and deterministic. The measurement equation is:

$$z(j) = C(j)\theta(j) + v(j), \quad j=1,\dots,k$$

in which $\{v(j)\}$ is a random sequence.

Define the following cost function

$$J(\theta(0)) = \sum_{j=1}^{k} \left(z(j) - C(j)\theta(j)\right)^T Q(j) \cdot \left(z(j) - C(j)\theta(j)\right)$$

where $Q(j) \in \mathbb{R}^{q \times q}$ are symmetric and positive definite (weights).

Our estimation problem is

$$\boxed{\tilde{\theta}(0|k) = \underset{\theta(0) \in \mathbb{R}^{n}}{\operatorname{argmin}}\ J(\theta(0))}$$

This is a static estimate of $\theta(0)$. If we let evolve $\tilde{\theta}(0|k)$ through the equation (12) we obtain the estimate $\tilde{\theta}(k|k)$ of $\theta(k)$ which we call (DYNAMIC) WEIGHTED LEAST SQUARES ESTIMATE OF $\theta(k)$

From (12)

$$\theta(j) = \phi(j,0)\,\theta(0) + h(j) \quad , j=1,\dots,k, \qquad (13)$$

where

$$\begin{cases} \phi(j,\tau) = A(j-1)A(j-2)\cdots A(\tau) \;, & j>\tau\geq 0 \\ \phi(\tau,\tau) = I_{\mu\times\mu} \;, & j=\tau \end{cases}$$

and

$$h(j) = \sum_{\tau=0}^{j-1} \phi(j,\tau+1)\, m(\tau)$$

Set

$$\phi^{(k)} = \begin{pmatrix} C(1)\,h(1) \\ \vdots \\ C(k)\,h(k) \end{pmatrix}, \quad H^{(k)} = \begin{pmatrix} C(1)\phi(1,0) \\ \vdots \\ C(k)\phi(k,0) \end{pmatrix}$$

then from (13)

$$\begin{pmatrix} z(1) - \mathcal{C}(1)\theta(1) \\ \vdots \\ z(k) - \mathcal{C}(k)\theta(k) \end{pmatrix} = z^{(k)} - H^{(k)}\theta(0) - p^{(k)}$$

If we also set $\mathcal{Q}^{[k]} = \mathrm{diag}\{\mathcal{Q}(1), \dots, \mathcal{Q}(k)\}$ as usual, then we obtain

$$J(\theta(0)) = \left(z^{(k)} - H^{(k)}\theta(0) - p^{(k)}\right)^T \mathcal{Q}^{[k]} \cdot$$
$$\cdot \left(z^{(k)} - H^{(k)}\theta(0) - p^{(k)}\right)$$
$$= \left(\tilde{z}^{(k)} - \tilde{\mathcal{C}}^{(k)}\theta(0)\right)^T \mathcal{Q}^{[k]} \left(\tilde{z}^{(k)} - \tilde{\mathcal{C}}^{(k)}\theta(0)\right)$$

where $\tilde{z}^{(k)} = z^{(k)} - p^{k}$, $\tilde{\mathcal{C}}^{(k)} = H^{(k)}$. Then the $J(\theta(0))$ is formally equal to the cost function $J(\theta)$ in the static estimation problem. Therefore, if we assume (in analogy) that

$$\boxed{\mathrm{rank}\ H^{(k)} = n} \qquad (14)$$

we obtain

$$\boxed{\tilde{\theta}(0|k)} =$$

$$= \left(\tilde{C}^{(k)T} Q^{[k]} \tilde{C}^{(k)}\right)^{-1} \tilde{C}^{(k)T} Q^{[k]} \tilde{z}^{(k)}$$

$$\boxed{= \left(H^{(k)T} Q^{[k]} H^{(k)}\right)^{-1} H^{(k)T} Q^{[k]} \left(z^{(k)} - p^{(k)}\right)}$$

Finally

$$\boxed{\tilde{\theta}(k|k) = \phi(k,0)\tilde{\theta}(0|k) + h(k)} \qquad (15)$$

As to the estimation error covariance, first from (13)

$$\tilde{e}(k|k) := \theta(k) - \tilde{\theta}(k|k) =$$

$$= \phi(k,0)\,\theta(0) + h(k) - \phi(k,0)\tilde{\theta}(0|k) - h(k)$$

$$= \phi(k,0)\left(\theta(0) - \tilde{\theta}(0|k)\right) = \phi(k,0)\,\tilde{e}(0|k) \qquad (16)$$

where $\hat{e}(0|k) := \theta(0) - \tilde{\theta}(0|k)$. Note also that $\tilde{\theta}(k|k)$ is centered by (15) and since $\tilde{\theta}(0|k)$ is centered (it is a least squares estimate)

Indeed,

$$E[\tilde{\theta}(k|k)] = \phi(k,0)E[\tilde{\theta}(0|k)] + h(k)$$
$$= \phi(k,0)\theta(0) + h(k) = \theta(k)$$

Therefore, from (16)

$$\Psi_{\tilde{e}(k|k)} = \Psi_{\tilde{\theta}(k|k)} =$$
$$= E\left[(\theta(k) - \tilde{\theta}(k|k))(\theta(k) - \tilde{\theta}(k|k))^T \right]$$
$$= E\left[\phi(k,0)\tilde{e}(0|k)\tilde{e}^T(0|k)\phi^T(k,0) \right]$$
$$= \phi(k,0)\Psi_{\tilde{e}(0|k)}\phi^T(k,0) \qquad (17)$$

Finally, since $\tilde{\theta}(0|k)$ is a least squares estimate

$$\Psi_{\tilde{e}(0|k)} = \Psi_{\tilde{\theta}(0|k)} =$$
$$= \left(\tilde{C}^{(k)^T} Q^{[k]} \tilde{C}^{(k)}\right)^{-1} \tilde{C}^{(k)} Q^{[k]} \Psi_{V(k)} Q^{[k]} \tilde{C}^{(k)} \left(\tilde{C}^{(k)^T} Q^{[k]} \tilde{C}^{(k)}\right)^{-1}$$
$$= \left(H^{(k)^T} Q^{[k]} H^{(k)}\right)^{-1} H^{(k)} Q^{[k]} \Psi_{V(k)} Q^{[k]} H^{(k)} \left(H^{(k)^T} Q^{[k]} H^{(k)}\right)^{-1}$$

which replaced in (17) gives $\Psi_{\tilde{e}(k|k)}$.

Assuming, in addition, that $\{v(j)\}$ is uncorrelated with nonsingular $\Psi_{v(j)}$, $j=1,\dots,k$, we have by setting $\boxed{Q^{[k]} = \Psi_{v(k)}^{-1}}$:

$$\tilde{\Theta}(k|k) = \Phi(k,0)\left(H^{(k)T}\Psi_{v(k)}^{-1}H^{(k)}\right)^{-1}H^{(k)T}\Psi_{v(k)}^{-1}\left(z^{(k)} - p^{(k)}\right) + h(k) \qquad (18)$$

with

$$\Psi_{\tilde{\Theta}(k|k)} = \Psi_{\tilde{\theta}(k|k)} = \Phi(k,0)\left(H^{(k)T}\Psi_{v(k)}^{-1}H^{(k)}\right)^{-1}\Phi^{T}(k,0)$$

(18) is called DYNAMIC MARKOV estimate of $\Theta(k)$.

MAXIMUM LIKELIHOOD ESTIMATE

Let $\theta \in R^{\mu}$ be deterministic and define the cost function

$$J(\theta) = p_{\underline{z}(k)}(\breve{z}^{(k)}, \theta) .$$

Whenever it exists,

$$\tilde{\theta}|_k = \underset{\theta \in \mathbb{R}^{\mu}}{\arg\max} \; J(\theta)$$

we say it is a MAXIMUM LIKELIHOOD estimate (ML). It is the choice which renders the measurements as likely to happen as possible. Since the logarithmic function is monotone increasing, we can consider equivalently the following cost

$$\boxed{J'(\theta) = \ln p_{z(k)}(\breve{z}^{(k)}, \theta)}$$

and again

$$\tilde{\theta}|_k = \underset{\theta}{\arg\max}\, J'(\theta) = \underset{\theta}{\arg\max}\, J(\theta)$$

A necessary condition for the existence of $\tilde{\theta}|_k$ is

$$\boxed{\frac{\partial}{\partial \theta} p_{z(k)}(\breve{z}^{(k)}, \theta)\Big|_{\theta = \tilde{\theta}|_k} = 0} \qquad (19)$$

or equivalently

$$\frac{\partial}{\partial \theta} \ln p_{z(k)}(\breve{z}^{(k)}, \theta)\Big|_{\theta = \tilde{\theta}|_k} = 0$$

PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATES

If

- $\{v(j)\}$ is independent and i.i.d
- $g(\theta, j) = g(\theta) \ \forall j = 1, .., k$
  and g is invertible (identifiability with one measurement)
- $\frac{\partial^i p_z(\zeta, \theta)}{\partial \theta^i}$, $i = 1, 2, 3$ ($p_z(\zeta, \theta)$ is the common density for all the measurements $z(1) .., z(k)$) is well-defined, $\forall \zeta \in \Omega_z$ $\forall \theta \in S(\theta^*, \rho)$, an open sphere in $\mathbb{R}^\mu$ around the true value θ^* of θ with radius $\rho > 0$
- $\left\| \frac{\partial^i p_z(\zeta, \theta)}{\partial \theta^i} \right\| \leq \varphi_i(\zeta, \theta)$, $i = 1, 2$ $\forall \zeta \in \Omega_z$ $\forall \theta \in S(\theta^*, \rho)$, φ_i integrabile with respect to $\zeta \in \Omega_z$ uniformly with respect to θ

- $\left\| E\left\{ \frac{\partial^3}{\partial\theta^3} \ln p_z(z,\theta) \right\} \right\| \leq M$

  $\forall\, \theta \in S(\theta^*, e)$

- $E\left\{ \left\| \frac{\partial}{\partial\theta} \ln p_z(z,\theta) \right\|^2 \right\} \in [\alpha, \beta]$,

  $\alpha > 0$, $\forall \theta \in S(\theta^*, e)$

---

then there exists a maximum likelihood estimate of θ for k sufficiently large, it is consistent, asymptotically gaussian, asymptotically centered and asymptotically efficient.

Here asymptotically means for $k \to \infty$

Other existence conditions are the following ones.

If

- $g^{(k)}$ is invertible ~~for~~ sufficientely large k (not necessarily we have $g(\theta,j) \equiv g(\theta)$)
- $\dfrac{\partial^i g(\theta,j)}{\partial\theta^i}$, $i=1,2,3$, are well defined $\forall\, \theta \in S(\theta^*, e)$
- $\left\| \dfrac{\partial^i}{\partial\theta^i} g(\theta,j) \right\| \leq k$, $i=1,2,3$ $\forall \theta \in S(\theta^*, e)$

  $k>0$
- $\| g(\theta^*, j) \| \leq k'$, $j=1,2$, $k'>0$
- $\lim\limits_{k\to\infty} \dfrac{1}{k} \sum\limits_{j=1}^{k} \left\| \dfrac{\partial g(\theta,j)}{\partial\theta}\Big|_{\theta=\theta^*} \right\|^2 \geq \beta$,

  $\beta > 0$

then the same conclusions as before hold true.

There is one case in which a maximum likelihood estimate has a specific structure and it is centered and efficient.

Assume that the conditions for the existence of estimate (6) are satisfied. Then, whenever a maximum likelyhood estimate exists, it coincides with (6) and, as such, it is the unique centered and efficient estimate.

Indeed, if condition for (6) are met, the function

$$\gamma(S^{(k)}, \theta) := \theta + R_k^{-1}(\theta) \left[\frac{\partial}{\partial \theta} \ln p_{s^{(k)}}(s^{(k)}, \theta) \right]^T$$

is independent of θ. Therefore

$$\gamma(S^{(k)}, \theta) = \gamma(S^{(k)}, \hat{\theta}|_k)$$

where $\hat{\theta}|_k$ is a maximum likelihood estimate (whenever it exists).

We have by the property of $\hat{\theta}|_k$ (see (19))

$$0 = \frac{\partial}{\partial\theta} \ln p_{z(k)}\left(z^{L(k)}, \theta\right)\Big|_{\theta=\hat{\theta}|_k}$$

$$= R_k(\hat{\theta}|_k)(\tilde{\theta}|_k - \hat{\theta}|_k)$$

which implies $\tilde{\theta} = \hat{\theta}|_k$ by invertibility of $R_k(\theta)$. ◁

An important tool for obtain maximum likelihood estimate is the INVARIANCE PROPERTY.

Assume $\theta \in \mathbb{R}^{\mu}$, $\eta \in \mathbb{R}^{\beta}$ and

$$\eta = f(\theta), \quad \mu \geq \beta$$

with $D_\eta \subset \mathbb{R}^{\beta}$, $D_\theta \subset \mathbb{R}^{\mu}$ their admissibility sets.

Assume also

$$D_\eta = f(D_\theta) \qquad (20)$$

in the sense that $\forall \theta \in D_\theta \ \exists \eta \in D_\eta$ such that $\bar{\eta} = f(\bar{\theta})$.

If $\hat{\theta}|_k$ is a maximum likelihood estimate of θ, then

$$\hat{\eta}|_k = f(\hat{\theta}|_k)$$

is a maximum likelihood estimate of η. Conversely, if $\hat{\eta}|_k$ is a maximum likelihood estimate of η then all the estimates

$$\hat{\theta}|_k \in \hat{\Theta} := \left\{ \hat{\theta} \in D_\theta : \hat{\eta}|_k = f(\hat{\theta}) \right\}$$

are maximum likelihood estimates.

Proof. Define the cost functions

$$J(\theta) := \ln p_{z(k)}(\tilde{z}^{(k)}, \theta)$$

where $p_{z(k)}(z^{(k)}, \theta) = p_{v(k)}(z^{(k)} - g^{(k)}(\theta))$

and

$$\tilde{L}(\eta) = \ln p_{z(k)}(\check{z}^{(k)}, \eta)$$

where $p_{z(k)}(z^{(k)}, \eta) = p_{v(k)}(z^{(k)} - h^{(k)}(\eta))$ and $h^{(k)}(\eta)$ is such that

$$z^{(k)} = g^{(k)}(\theta) + v^{(k)} = h^{(k)}(\eta) + v^{(k)}$$

(equivalently $h^{(k)}(f(\theta)) = g^{(k)}(\theta))$.

In other words, J and L are the same cost function, the first as a function of θ, the second as a function of η. Therefore

$$(21) \quad \boxed{J(\theta) = L(\eta) \quad \text{whenever } \eta = f(\theta)}$$

Let $\hat{\theta}|_k \in D_\theta$ be a maximum likelihood estimate of θ and assume that

$$\hat{\eta}|_k = f(\hat{\theta}|_k) \in D_\eta$$

is not a maximum likelihood estimate of η. There exists $\bar{\eta} \in D_\eta$, $\bar{\eta} \neq \hat{\eta}|_k$ such that

$$L(\bar{\eta}) > L(\hat{\eta}|_k)$$

By (20) it follows the existence of $\bar{\theta} \in D_\theta$ such that

$$\bar{\eta} = f(\bar{\theta}) \quad , \quad \bar{\theta} \neq \hat{\theta}|_k$$

and, consequently, using (21)

$$J(\bar{\theta}) = L(\bar{\eta}) > L(\hat{\eta}|_k) = J(\hat{\theta}|_k)$$

This gives a contradiction since $\hat{\theta}|_k$ maximizes $J(\theta)$.

Conversely, let $\hat{\eta}|_k \in D_\eta$ a maximum likelihood estimate of η. One has by (21)

$$J(\hat{\theta}|_k) = L(\hat{\eta}|_k) \quad \forall \hat{\theta}|_k \in \hat{\Theta}$$

(i.e. $\forall \hat{\theta}|_k$ such that $\hat{\eta}|_k = f(\hat{\theta}|_k)$).

Assume the existence of $\bar{\theta} \in D_\theta \setminus \hat{\Theta}$ such that

$$J(\bar{\theta}) > J(\hat{\theta}|_k) \quad \forall \hat{\theta}|_k \in \hat{\Theta}$$

and $\bar{\eta} = f(\bar{\theta}) \in D_\eta$, $\bar{\eta} \neq \hat{\eta}|_k$. It would follow by (21)

$$L(\bar{\eta}) = J(\bar{\theta}) > J(\hat{\theta}|_k) = L(\hat{\eta}|_k)$$

which gives a contradiction, since $\hat{\eta}|_k$ maximizes $L(\eta)$. $\triangleleft$

An important case in which the maximum likelihood estimation can be seen as a "nonlinear" extension of weighted least squares estimation is the one of gaussian noise sequence $\{v(j)\}$. Assume $E[v(j)] = 0 \quad \forall j = 1, \dots, k$. We have

$$p_{z(k)}(\zeta^{(k)}, \theta) = \frac{1}{(2\pi)^{kq/2}\left(\det \Psi_{v(k)}\right)^{1/2}} \cdot e^{-\frac{1}{2}\left(\zeta^{(k)} - g^{(k)}(\theta)\right)^T \Psi_{v(k)}^{-1} \left(\zeta^{(k)} - g^{(k)}(\theta)\right)}$$

then, if $J(\theta) := \ln p_{z(k)}(\breve{z}^{(k)}, \theta)$,

$$J(\theta) = -\ln\left[(2\pi)^{kq/2}\left(\det \Psi_{v(k)}\right)^{1/2}\right] - \frac{1}{2}\left[\breve{z}^{(k)} - g^{(k)}(\theta)\right]^T \Psi_{v(k)}^{-1} \left[\breve{z}^{(k)} - g^{(k)}(\theta)\right]$$

and

$$\arg\max_{\theta \in D_\theta} J(\theta) = \arg\max_{\theta \in D_\theta} J'(\theta)$$

$$= \operatorname*{argmin}_{\theta \in D_\theta} J''(\theta)$$

where $J'(\theta) := [\check{z}^{(k)} - g^{(k)}(\theta)]^T \Psi_{\nu(k)}^{-1} [\check{z}^{(k)} - g^{(k)}(\theta)]$

$$J''(\theta) := - J'(\theta)$$

By the necessary condition (19)

$$0 = \frac{dJ''}{d\theta}\bigg|_{\theta=\hat{\theta}|_k} = -2[\check{z}^{(k)} - g^{(k)}(\theta)] \Psi_{\nu(k)}^{-1} \frac{dg^{(k)}}{d\theta}\bigg|_{\theta=\hat{\theta}|_k}$$

If $g^{(k)}(\theta) = C^{(k)}\theta$ for some $C^{(k)}$ such that $\operatorname{rank} C^{(k)} = \mu$, we recover the weighted least squares estimates.

Under this regard, we remark that the invariance property can be used together with a re-parametrization of the estimation problem.

For example, say that

$$g^{(k)}(\theta) = C^{(k)} f(\theta) = C^{(k)} \eta$$

for some $C^{(k)}$ and f such that:

- $\operatorname{rank} C^{(k)} = \mu$
- $D_\eta = f(D_\theta)$

Therefore, a least squares estimate for η is given by

$$\hat{\eta}|_k = \left(C^{(k)T} \Psi_{V(k)}^{-1} C^{(k)} \right)^{-1} C^{(k)T} \Psi_{V(k)}^{-1} \breve{z}^{(k)}$$

By the invariance property any $\hat{\theta}|_k$ such that

$$\hat{\eta}|_k = f(\hat{\theta}|_k)$$

is a maximum likelihood estimate of η, whenever $\hat{\eta}|_k$ is also a maximum likelihood estimates, for example when $\{v(j)\}$ is gaussian, as we have shown above.

Some concluding remarks on the dynamic maximum likelihood estimation problem. In this case,

$$\begin{cases} \theta(j+1) = f(\theta(j), m(j)), & j = 1, \ldots, k-1 \\ z(j) = g(\theta(j), j) + v(j), & j = 1, \ldots, k \end{cases}$$

In order to find a maximum likelihood estimate of $\theta(k)$, we first obtain for $j = 1, \ldots, k$

$$\theta(j) = \varphi(\theta(0), m(0), \ldots, m(j-1), j) \quad (22)$$

for some function $\varphi(\cdot, j)$ and then

$$g(\theta(j), j) = h(\theta(0), m(0), \ldots, m(j-1), j)$$

for some function $h(\cdot, j)$ (we are assuming $\{m(j)\}$ is deterministic and known).

A maximum likelihood estimate for $\theta(0)$ can be sought and from this we obtain a maximum likelihood estimate for $\theta(k)$ from the parameter equation (22)

$$\hat{\theta}(k|k) = \varphi(\hat{\theta}(0|k), m(0), \ldots, m(k-1), k)$$

BAYESIAN ESTIMATES

Let $\mathcal{C} : \mathbb{R}^{\mu} \to \mathbb{R}$ be a non-negative function and define the cost function

$$J(\alpha) := E\{ \mathcal{C}(\theta - \alpha) | z^{(k)} \}$$
$$= \int_{\Omega_{\theta|z^{(k)}}} \mathcal{C}(\sigma - \alpha) \, p_{\theta|z^{(k)}}(z^{(k)}, \sigma) \, d\sigma$$

where θ is the random parameter vector, $\theta \in \mathbb{R}^{\mu}$. We define

$$\tilde{\theta}|_k := \operatorname*{argmin}_{\alpha \in \Omega_{\theta|z^{(k)}}} J(\alpha)$$

as the BAYESIAN ESTIMATE which minimizes J.

For evaluating J, it is necessary to evaluate first $p_{\Theta|Z^{(k)}}$.

This can be done as follows

$$p_{\Theta|Z^{(k)}}(\zeta^{(k)}, \sigma) = \frac{p_{\binom{\Theta}{Z^{(k)}}}(\zeta^{(k)}, \sigma)}{p_{Z^{(k)}}(\zeta^{(k)})}$$

N.B. by definition of conditional density (see (A.?));

$p_{\binom{\Theta}{Z^{(k)}}}$ is the joint density of Θ and $Z^{(k)}$

$$= \frac{p_{\binom{\Theta}{Z^{(k)}}}(\zeta^{(k)}, \sigma)}{\int_{\Omega_\Theta} p_{\binom{\Theta}{Z^{(k)}}}(\zeta^{(k)}, \eta)\, d\eta}$$

N.B by (A.3)

$$= \frac{p_{Z^{(k)}|\Theta}(\zeta^{(k)}, \sigma)\, p_\Theta(\sigma)}{\int_{\Omega_\Theta} p_{Z^{(k)}|\Theta}(\zeta^{(k)}, \eta)\, p_\Theta(\eta)\, d\eta}$$

by again (A.3)

We reduced the evaluation of $p_{\theta | z^{(k)}}$ to the knowledge of $p_{z^{(k)} | \theta}$ which can be evaluated from the a priori information p_θ and $p_{v^{(k)}}$. For example if

$$z^{(k)} = g^{(k)}(\theta) + v^{(k)}$$

then

$$\boxed{p_{z^{(k)} | \theta}(\zeta^{(k)}, \sigma) = p_{v^{(k)}}\left(\zeta^{(k)} - g^{(k)}(\theta)\right)}$$

In these notes we consider only one class of Bayesian estimates, the ones for which

$$\boxed{\begin{aligned} \mathcal{C}(\theta - \alpha) &:= \| \theta - \alpha \|^2 \\ &= (\theta - \alpha)^T (\theta - \alpha) \end{aligned}}$$

These Bayesian estimates are known as MINIMUM MEAN SQUARE ERROR (MMSE) estimates.

We want to prove some basic properties of these MMSE estimates.

First of all,

$\boxed{\text{1. A MMSE estimate } \tilde{\theta}|_k \text{ can be evaluated as } \tilde{\theta}|_k = E[\theta \mid z^{(k)}] \text{ as long as } \Omega_{\theta|z^{(k)}} \text{ is a convex set}}$

Proof. We have by definition of MMSE estimate $\tilde{\theta}|_k$

$$\tilde{\theta}|_k = \arg\max_{\alpha \in \Omega_{\theta|z^{(k)}}} J(\alpha)$$

where $J(\alpha) := E[\|\theta - \alpha\|^2 \mid z^{(k)}]$.

By necessary conditions,

$$\left.\frac{\partial J}{\partial \alpha}\right|_{\alpha = \tilde{\theta}|_k} = 0$$

Therefore,

$$0 = \left.\frac{\partial J}{\partial \alpha}\right|_{\alpha = \tilde{\theta}|_k}$$

$$= \left.\frac{\partial}{\partial \alpha} \int_{\Omega_{\Theta|z^{(k)}}} p_{\Theta|z^{(k)}}(z^{(k)}, \sigma) \, \|\sigma - \alpha\|^2 \, d\sigma \right|_{\alpha = \tilde{\theta}|_k}$$

$$= \left.\int_{\Omega_{\Theta|z^{(k)}}} p_{\Theta|z^{(k)}}(z^{(k)}, \sigma) \, \frac{\partial}{\partial \alpha} \|\sigma - \alpha\|^2 \, d\sigma \right|_{\alpha = \tilde{\theta}|_k}$$

$$= 2 \left.\int_{\Omega_{\Theta|z^{(k)}}} p_{\Theta|z^{(k)}}(z^{(k)}, \sigma) \, (\sigma - \alpha)^T \, d\sigma \right|_{\alpha = \tilde{\theta}|_k}$$

$$= 2 \int_{\Omega_{\Theta|z^{(k)}}} p_{\Theta|z^{(k)}}(z^{(k)}, \sigma) \, \sigma^T \, d\sigma$$

$$- 2 \int_{\Omega_{\Theta|Z^{(k)}}} p_{\Theta|Z^{(k)}}(z^{(k)}, \sigma) \tilde{\Theta}\Big|_k^T d\sigma$$

$$= 2E[\Theta^T | z^{(k)}] - 2\tilde{\Theta}\Big|_k^T \qquad (23)$$

(since $\int_{\Omega_{\Theta|Z^{(k)}}} p_{\Theta|Z^{(k)}}(z^{(k)}, \sigma) d\sigma = 1$).

We obtain $\tilde{\Theta}\Big|_k = E[\Theta | z^{(k)}]$
by taking transposes in (23).

Moreover,

$$\frac{\partial^2 J}{\partial \alpha^2} = \frac{\partial}{\partial \alpha}\left(\frac{\partial J}{\partial \alpha}\right)^T =$$

$$= 2\frac{\partial}{\partial \alpha} \int_{\Omega_{\Theta|Z^{(k)}}} (\sigma - \alpha) p_{\Theta|Z^{(k)}}(z^{(k)}, \sigma) d\sigma$$

$$= 2 \int_{\Omega_{\Theta|Z^{(k)}}} p_{\Theta|Z^{(k)}}(z^{(k)}, \sigma) d\sigma = 2 I_{\mu \times \mu}$$

It follows that J is strictly convex and $\tilde{\Theta}|_k$, whenever $\tilde{\Theta}|_k \in \Omega_{\Theta|z^{(k)}}$, is the sought minimum for J. But

$$\tilde{\Theta}|_k = E[\Theta | z^{(k)}] \in \mathrm{Convclos}(\Omega_{\Theta|z^{(k)}})$$

where Convclos($\mathcal{S}$) is the convex closure of the set $\mathcal{S}$, i.e. $\mathrm{Convclos}(\mathcal{S}) := \{ s \in \mathbb{R}^{\mu} : s = \gamma a + b(1-\gamma), \gamma \in \mathbb{R}, a, b \in \mathbb{R}^{\mu} \}$.

Since $\Omega_{\Theta|z^{(k)}}$ is convex by assumption,

$$\mathrm{Convclos}(\Omega_{\Theta|z^{(k)}}) = \Omega_{\Theta|z^{(k)}}$$

and $\tilde{\Theta}|_k = E[\Theta | z^{(k)}]$ is indeed the only admissible solution. $\triangleleft$

N.B. $\tilde{\Theta}|_k = E[\Theta | z^{(k)}]$ is always centered since

$E[\tilde{\Theta}|_k] = E[E[\Theta | z^{(k)}]] = E[\Theta]$.

2. If $p_{\binom{\theta}{z^{(k)}}}$ is gaussian, then the MMSE estimate $\hat{\theta}|_k$ is affine with respect to $z^{(k)}$

Proof. We use the notations

$$E\left[\begin{pmatrix}\theta\\ z^{(k)}\end{pmatrix}\right] = \begin{pmatrix}\bar{\theta}\\ \overline{z^{(k)}}\end{pmatrix}$$

$$\Psi_{\binom{\theta}{z^{(k)}}} = E\left[\begin{pmatrix}\theta-\bar{\theta}\\ z^{(k)}-\overline{z^{(k)}}\end{pmatrix}\begin{pmatrix}\theta-\bar{\theta}\\ z^{(k)}-\overline{z^{(k)}}\end{pmatrix}^T\right]$$

$$= \begin{pmatrix}\Psi_{\theta} & \Psi_{\theta z^{(k)}}\\ \Psi_{z^{(k)}\theta} & \Psi_{z^{(k)}}\end{pmatrix}$$

First

$$\Psi^{-1}_{\binom{\theta}{z^{(k)}}} = \begin{pmatrix} \Psi_\theta & \Psi_{\theta z^{(k)}} \\ \Psi_{z^{(k)}\theta} & \Psi_{z^{(k)}} \end{pmatrix}^{-1} \qquad (24)$$

$$= \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \qquad \left(\begin{array}{l} \text{N.B. remember} \\ \Psi_{z^{(k)}\theta} = \Psi_{\theta z^{(k)}}^T \end{array} \right)$$

where $A := \left(\Psi_\theta - \Psi_{\theta z^{(k)}} \Psi^{-1}_{z^{(k)}} \Psi_{z^{(k)}\theta}\right)^{-1}$

$B := -A \Psi_{\theta z^{(k)}} \Psi^{-1}_{z^{(k)}}$

$C := \Psi^{-1}_{z^{(k)}} + \Psi^{-1}_{z^{(k)}} \Psi_{z^{(k)}\theta} A \Psi_{\theta z^{(k)}} \Psi^{-1}_{z^{(k)}}$

N.B. The inverse matrix $\Psi^{-1}_{\binom{\theta}{z^{(k)}}}$ exists since $\binom{\theta}{z^{(k)}}$ is gaussian, similarly Ψ^{-1}_θ and $\Psi^{-1}_{z^{(k)}}$ exist since θ and $z^{(k)}$ are marginally gaussian as subvectors of $\binom{\theta}{z^{(k)}}$.

Indeed, if M is nonsingular and

$$M := \begin{pmatrix} E & F \\ F^T & G \end{pmatrix}$$

then, using the so-called Schur complement

$$M^{-1} = \begin{pmatrix} E & F \\ F^T & G \end{pmatrix}^{-1} =$$

$$= \left[\begin{array}{c|c} (E - FG^{-1}F^T)^{-1} & -(E-FG^{-1}F^T)^{-1}FG^{-1} \\ \hline -G^{-1}F^T(E-FG^{-1}F^T)^{-1} & G^{-1}+G^{-1}F^T(E-FG^{-1}F^T)^{-1}FG^{-1} \end{array}\right]$$

which leads to (24). Next, using the fact that $p_{\binom{\theta}{z(k)}}$ and, marginally, $p_{z(k)}$ are both gaussian and from (24)

$$p_{z^{(k)}}(\varsigma^{(k)}) = \frac{1}{(2\pi)^{\frac{kq}{2}}} \cdot \frac{1}{\det \Psi_{z^{(k)}}} \cdot e^{-\frac{1}{2}(\varsigma^{(k)} - \bar{z}^{(k)})^T \Psi_{z^{(k)}}^{-1} (\varsigma^{(k)} - \bar{z}^{(k)})}$$

$$p_{\binom{\theta}{z^{(k)}}}(\sigma, \varsigma^{(k)}) = \frac{1}{(2\pi)^{\frac{kq+\mu}{2}}} \cdot \frac{1}{\det \Psi_{\binom{\theta}{z^{(k)}}}} \cdot e^{-\frac{1}{2}\begin{pmatrix}\sigma - \bar{\theta} \\ \varsigma^{(k)} - \bar{z}^{(k)}\end{pmatrix}^T \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix}\sigma - \bar{\theta} \\ \varsigma^{(k)} - \bar{z}^{(k)}\end{pmatrix}}$$

therefore using the definition of conditional density

$$p_{\theta|z^{(k)}}(\sigma, \varsigma^{(k)}) = \frac{p_{\binom{\theta}{z^{(k)}}}(\sigma, \varsigma^{(k)})}{p_{z^{(k)}}(\varsigma^{(k)})} =$$

$$= \frac{1}{(2\pi)^{\mu/2}} \cdot \frac{\det \Psi_{z^{(k)}}}{\det \Psi_{\binom{\theta}{z^{(k)}}}} \cdot e^{-\frac{1}{2}\left\{\begin{pmatrix}\sigma - \bar{\theta} \\ \varsigma^{(k)} - \bar{z}^{(k)}\end{pmatrix}^T \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix}\sigma - \bar{\theta} \\ \varsigma^{(k)} - \bar{z}^{(k)}\end{pmatrix} - (\varsigma^{(k)} - \bar{z}^{(k)})^T \Psi_{z^{(k)}}^{-1} (\varsigma^{(k)} - \bar{z}^{(k)})\right\}}.$$

Next, as a consequence of

$$\left(\begin{array}{c|c} I & -\Psi_{\theta z^{(k)}} \Psi_{z^{(k)}}^{-1} \\ \hline O & I \end{array}\right) \Psi_{\binom{\theta}{z^{(k)}}} \left(\begin{array}{c|c} I & -\Psi_{\theta z^{(k)}} \Psi_{z^{(k)}}^{-1} \\ \hline O & I \end{array}\right)^T =$$

$$= \left(\begin{array}{c|c} A^{-1} & O \\ \hline O & \Psi_{z(k)} \end{array}\right)$$

and

$$\det \left(\begin{array}{c|c} I & -\Psi_{\theta z(k)} \Psi_{z(k)}^{-1} \\ \hline O & I \end{array}\right) =$$

$$\det\left[\left(\begin{array}{c|c} I & -\Psi_{\theta z(k)} \Psi_{z(k)}^{-1} \\ \hline O & I \end{array}\right)^{T}\right] = 1$$

with $\det \left(\begin{array}{c|c} A^{-1} & O \\ \hline O & \Psi_{z(k)} \end{array}\right) = \det A^{-1} \cdot \det \Psi_{z(k)}$

we have $\det \Psi_{\binom{\theta}{z(k)}} = \det \Psi_{z(k)} \cdot \det A^{-1}$ and

$$\frac{\det \Psi_{z(k)}}{\det \Psi_{\binom{\theta}{z(k)}}} = \frac{1}{\det A^{-1}}$$

Finally,

$$\begin{pmatrix}\sigma-\bar\theta\\ \zeta^{(k)}-\bar z^{(k)}\end{pmatrix}^T \begin{pmatrix}A & B\\ B^T & C\end{pmatrix}\begin{pmatrix}\sigma-\bar\theta\\ \zeta^{(k)}-\bar z^{(k)}\end{pmatrix} - (\zeta^{(k)}-\bar z^{(k)})^T \Psi_{z^{(k)}}^{-1}(\zeta^{(k)}-\bar z^{(k)})$$

$$= (\sigma-\bar\theta)^T A(\theta-\bar\theta) + 2(\sigma-\bar\theta)^T B(\zeta^{(k)}-\bar z^{(k)})$$
$$+ (\zeta^{(k)}-\bar z^{(k)})^T (C-\Psi_{z^{(k)}}^{-1})(\zeta^{(k)}-\bar z^{(k)})$$

$$= (\sigma-\bar\theta)^T A(\theta-\bar\theta) - 2(\sigma-\bar\theta)^T A\Psi_{\theta z^{(k)}}\Psi_{z^{(k)}}^{-1}(\zeta^{(k)}-\bar z^{(k)})$$
$$+ (\zeta^{(k)}-\bar z^{(k)})^T \Psi_{z^{(k)}}^{-1}\Psi_{z^{(k)}\theta} A\Psi_{\theta z^{(k)}}\Psi_{z^{(k)}}^{-1}(\zeta^{(k)}-\bar z^{(k)})$$

$$= \left(\sigma-\bar\theta-\Psi_{\theta z^{(k)}}\Psi_{z^{(k)}}^{-1}(\zeta^{(k)}-\bar z^{(k)})\right)^T \cdot A \cdot$$
$$\cdot \left(\sigma-\bar\theta-\Psi_{\theta z^{(k)}}\Psi_{z^{(k)}}^{-1}(\zeta^{(k)}-\bar z^{(k)})\right)$$

We conclude that

$$p_{\theta|z^{(k)}}(\sigma,\zeta^{(k)}) = \frac{1}{(2\pi)^{n/2}}\cdot\frac{1}{(\det A^{-1})}\cdot$$
$$e^{-\frac{1}{2}\left(\sigma-\bar\theta-\Psi_{\theta z^{(k)}}\Psi_{z^{(k)}}^{-1}(\zeta^{(k)}-\bar z^{(k)})\right)^T A\left(\sigma-\bar\theta-\Psi_{\theta z^{(k)}}\Psi_{z^{(k)}}^{-1}(\zeta^{(k)}-\bar z^{(k)})\right)}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}}\cdot\frac{1}{(\det\Psi_{\theta|z^{(k)}})}\cdot e^{-\frac{1}{2}\left[(\sigma-E[\theta|z^{(k)}])^T\Psi_{\theta|z^{(k)}}^{-1}(\sigma-E[\theta|z^{(k)}])\right]}$$

which means that $p_{\theta|z^{(k)}}$ is gaussian

(91)

with

$$\boxed{\begin{aligned} \Psi_{\theta|z^{(k)}} &= A^{-1} = \Psi_{\theta} - \Psi_{\theta z^{(k)}} \Psi_{z^{(k)}}^{-1} \Psi_{z^{(k)}\theta} \\ E\left[\theta|z^{(k)}\right] &= \bar{\theta} + \Psi_{\theta z^{(k)}} \Psi_{z^{(k)}}^{-1} \left(z^{(k)} - \bar{z}^{(k)}\right) \end{aligned}} \tag{25}$$

and this proves also that the MMSE estimate $\tilde{\theta}|_k = E\left[\theta/z^{(k)}\right]$ is affine with respect to $z^{(k)}$. $\triangleleft$

When given $p_{v^{(k)}}$ and p_θ gaussian, also $p_{\theta|z^{(k)}}$ is gaussian? We want to discuss this issue here. We need basic results on gaussian variables. Let $X, Y \in \mathbb{R}$ gaussian and uncorrelated. Consider any linear combination

$$aX + bY, \quad a, b \in \mathbb{R}.$$

First of all, if X is gaussian with mean $\overline{X}$ and variance σ_X^2 then aX is gaussian with mean $a\overline{X}$ and variance $a^2\sigma_X^2$.

Indeed, by (A.1), if $a \neq 0$

$$p_{aX}(x) = p_X\left(\frac{x}{a}\right)\frac{1}{|a|}$$

$$= \frac{1}{(2\pi)^{1/2}} \cdot \frac{1}{|a|\sigma_X} \cdot e^{-\frac{1}{2\sigma_X^2}\left(\frac{x}{a}-\overline{X}\right)^2}$$

$$= \frac{1}{(2\pi)^{1/2}}\,\frac{1}{|a|\sigma_X}\,e^{-\frac{1}{2}\cdot\frac{1}{a^2\sigma_X^2}(x-a\overline{X})^2}.$$

Next, consider the invertible transformation

$$\begin{pmatrix} P \\ Q \end{pmatrix} = f\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X+Y \\ Y \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$$

where X, Y are gaussian and uncorrelated.

It follows again from (A.?) that

$$p_{\binom{P}{Q}}(p,q) = p_{\binom{X}{Y}}\left(f^{-1}(p,q)\right)\left|\det \frac{\partial f^{-1}}{\partial(p,q)}(p,q)\right|$$

But X, Y, since gaussian and uncorrelated are also independent:

$$p_{\binom{X}{Y}}(x,y) = p_X(x)\, p_Y(y)$$

Therefore, since $f^{-1}(p,q) = (p-q,q)$

$$p_{\binom{P}{Q}}(p,q) = p_X(p-q)\, p_Y(q)\, \frac{1}{|a|}.$$

If $\bar{X} := E[X]$ and $\bar{Y} := E[Y]$,

then

$$p_P(p) = \int_{\Omega_Q} p_{\binom{P}{Q}}(p,q)\, dq$$

$$= \int_{\mathbb{R}} p_X(p-q)\, p_Y(q)\, dq$$

$$= \int_{\mathbb{R}} \frac{1}{(2\pi)^{1/2}} \frac{1}{\sigma_X} e^{-\frac{1}{2\sigma_X^2}(p-q-\bar{X})^2} \cdot$$

$$\cdot \frac{1}{(2\pi)^{1/2}} \frac{1}{\sigma_Y} e^{-\frac{1}{2\sigma_Y^2}(q-\bar{Y})^2} dq$$

$$= \int_{\mathbb{R}} \frac{1}{(2\pi)^{1/2}\sqrt{\sigma_X^2+\sigma_Y^2}} \cdot e^{-\frac{(p-(\bar{X}+\bar{Y}))^2}{2(\sigma_X^2+\sigma_Y^2)}} \cdot$$

$$\cdot \frac{1}{(2\pi)^{1/2} \frac{\sigma_X \sigma_Y}{\sqrt{\sigma_X^2+\sigma_Y^2}}} e^{-\frac{\left(q-\frac{\sigma_Y^2(p-\bar{X})+\sigma_X^2\bar{Y}}{\sigma_X^2+\sigma_Y^2}\right)^2}{2\left(\frac{\sigma_X\sigma_Y}{\sqrt{\sigma_X^2+\sigma_Y^2}}\right)^2}} dq$$

$$= \frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}} e^{-\frac{(p-(\bar{X}+\bar{Y}))^2}{2(\sigma_X^2+\sigma_Y^2)}} \cdot$$

$$\cdot \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\, \frac{\sigma_X\sigma_Y}{\sigma_X^2+\sigma_Y^2}} e^{-\frac{\left(q-\frac{\sigma_Y^2(p-\bar{X})+\sigma_X^2\bar{Y}}{\sigma_X^2+\sigma_Y^2}\right)^2}{2\left(\frac{\sigma_X\sigma_Y}{\sqrt{\sigma_X^2+\sigma_Y^2}}\right)^2}} dq}_{(=1)}$$

$$\boxed{= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2+\sigma_Y^2}} \cdot e^{-\frac{(p-(X+Y))^2}{2(\sigma_X^2+\sigma_Y^2)}}}$$

which proves that $X+Y$ is gaussian with mean $\overline{X}+\overline{Y}$ and variance $\sigma_X^2+\sigma_Y^2$. Therefore, $aX+bY$ is gaussian for all $a,b\in\mathbb{R}$ with mean $a\overline{X}+b\overline{Y}$ and variance $a^2\sigma_X^2+b^2\sigma_Y^2$. We conclude

> 2.1. $X,Y\in\mathbb{R}$, gaussian and uncorrelated $\Rightarrow aX+bY$ gaussian for all $a,b\in\mathbb{R}$

Next, by definition of gaussian vector $\begin{pmatrix}X\\Y\end{pmatrix}$:

> 2.2. $\begin{pmatrix}X\\Y\end{pmatrix}$, with $X,Y\in\mathbb{R}$, is gaussian if and only if $aX+bY$ is gaussian for all $a,b\in\mathbb{R}$

Consider

$$z(j) = C(j)\theta + v(j), \; j = 1, .., k$$

where $\theta \in \mathbb{R}^{\mu}$, $\{v(j)\}$ are gaussian and $\{v(j)\}$ is white and mutually uncorrelated with θ. By 2.2. $C(j)\theta$ is gaussian and mutually uncorrelated with $\{v(j)\}$. Since any linear combination of $z(1), .., z(k)$ is a linear combination of θ and $\{v(j)\}$ and, moreover, θ and $\{v(j)\}$ are jointly gaussian (because they are independent and gaussian), then $z^{(k)}$ is gaussian. Moreover, by 2.2. again, $\begin{pmatrix} \theta \\ z^{(k)} \end{pmatrix}$ is gaussian since any linear combination of θ and $z^{(k)}$ is a linear combination of θ and $\{v(j)\}$.

If θ is time-varying and

$$\theta(j+1) = A(j)\theta(j) + B(j)m(j)$$
$$j = 0, \ldots, k-1,$$
$$z(j) = C(j)\theta(j) + v(j)$$

assuming that $\{v(j)\}$, $\{m(j)\}$ are gaussian, white, mutually uncorrelated and $\theta(0)$ is gaussian and mutually uncorrelated with $\{m(j)\}$ and $\{v(j)\}$, then by similar arguments we obtain that $\begin{pmatrix}\theta(i)\\ z(k)\end{pmatrix}$ is gaussian for all $i \geq 0$.

This establishes under which conditions $p_{\begin{pmatrix}\theta\\ z(k)\end{pmatrix}}$ (or $p_{\begin{pmatrix}\theta(i)\\ z(k)\end{pmatrix}}$ in the case of time-varying θ) is gaussian. It is clear it is needed that the measurement equation (and the parameter equation if θ is time-varying) is linear.

Now, we prove an optimality criterion for $\tilde{\theta}|_k$ being a MMSE estimate.

3. Let $\Psi_{z^{(k)}}$ be nonsingular and $\{\hat{\theta}|_k\}$ be the family of centered and affine with respect to $z^{(k)}$ estimates. $\tilde{\theta}|_k \in \{\hat{\theta}|_k\}$ is a MMSE estimate if and only if

$$E\left[\tilde{e}|_k \, z^{(k)T}\right] = 0 \qquad (25)$$

where $\tilde{e}|_k := \theta - \tilde{\theta}|_k$.

Proof. The family $\{\hat{\theta}|_k\}$ can be characterized as

$$\hat{\theta}|_k = \hat{\Lambda}^{(k)}\left(z^{(k)} - \bar{z}^{(k)}\right) + \bar{\theta}$$

where $\bar{z}^{(k)} := E[z^{(k)}]$ and $\bar{\theta} := E[\theta]$, where $\hat{\Lambda}^{(k)} \in \mathbb{R}^{\mu \times kq}$. Let

$$\tilde{\theta}|_k = \tilde{\Lambda}^{(k)}(z^{(k)} - \bar{z}^{(k)}) + \bar{\theta}$$

be a MMSE estimate (recall that this estimate is always centered). Therefore, we can parametrize $\hat{\Lambda}^{(k)}$ as follows

$$\hat{\Lambda}^{(k)} = \tilde{\Lambda}^{(k)} + \epsilon \Delta$$

where $\epsilon \in \mathbb{R}$ and $\Delta \in \mathbb{R}^{\mu \times qk}$.

Moreover, if $\hat{e}|_k := \theta - \hat{\theta}|_k$

$$E[\hat{e}|_k^T \hat{e}|_k | z^{(k)}] = E[(\tilde{e}|_k - \epsilon\Delta(z^{(k)} - \bar{z}^{(k)}))^T \cdot (\tilde{e}|_k - \epsilon\Delta(z^{(k)} - \bar{z}^{(k)})) | z^{(k)}]$$

and
$$J(\hat{\theta}|_k) = E[\|\hat{e}|_k\|^2 | z^{(k)}] \qquad (26)$$
$$= E[\hat{e}|_k^T \hat{e}|_k | z^{(k)}]$$

where J is the cost function minimized by $\tilde{\theta}|_k$.

Assume first that $\tilde{\theta}|_k$ is a MMSE estimate. It follows that (26) is minimum for $\epsilon = 0$ (since for this value $J(\hat{\theta}|_k) = J(\tilde{\theta}|_k)$):

$$0 = \frac{\partial}{\partial \epsilon} J(\hat{\theta}|_k)\Big|_{\epsilon=0} =$$

$$= -2E\left\{\left[\tilde{e}|_k - \epsilon\Delta(z^{(k)} - \bar{z}^{(k)})\right]^T \Delta(z^{(k)} - \bar{z}^{(k)}) \Big| z^{(k)}\right\}_{\epsilon=0}$$

$$= -2E\{\tilde{e}|_k^T \Delta(z^{(k)} - \bar{z}^{(k)}) \mid z^{(k)}\}$$

But applying expectation to both parts

$$0 = -2E\left[E\{\tilde{e}_k^T \Delta(z^{(k)} - \bar{z}^{(k)}) \mid z^{(k)}\}\right] =$$

$$= -2E\{\tilde{e}|_k^T \Delta(z^{(k)} - \bar{z}^{(k)})\} =$$

$$= -2E\{\tilde{e}|_k^T \Delta z^{(k)}\} + \underbrace{2E\{\tilde{e}|_k \Delta \bar{z}^{(k)}\}}_{=0}$$

since $\hat{\theta}|_k$ is centered

$$= -2E\{\tilde{e}|_k^T \Delta z^{(k)}\}$$

By arbitrariness of $\Delta \in \mathbb{R}^{\mu \times q k}$, we obtain (25) (note that (25) to uncorrelation of each component of $z^{(k)}$ with each comp. of $\tilde{e}|_k$).

Next, assume (25) holds true for some $\hat{\theta}|_k \in \{\theta|_k\}$. It follows

$$0 = E\{\hat{e}|_k z^{(k)T}\} = E\{[\tilde{e}|_k - \epsilon\Delta(z^{(k)} - \bar{z}^{(k)})] \cdot z^{(k)T}\} = -\epsilon\Delta E[(z^{(k)} - \bar{z}^{(k)}) z^{(k)T}]$$

N.B. since $E[\tilde{e}|_k z^{(k)T}] = 0$ being $\tilde{\theta}|_k$ a MMSE and we proved the necessity of (25) before

$$= -\epsilon\Delta E[(z^{(k)} - \bar{z}^{(k)})(z^{(k)} - \bar{z}^{(k)})^T] \underbrace{- \epsilon\Delta E[(z^{(k)} - \bar{z}^{(k)})\bar{z}^{(k)}]}_{(=0 \text{ since } E[z^{(k)} - \bar{z}^{(k)}] = 0)}$$

$$= -\epsilon\Delta \Psi_{z^{(k)}}.$$

Since $\Psi_{z^{(k)}}$ is nonsingular, $\epsilon\Delta = 0$ which implies $\hat{\theta}|_k = \tilde{\theta}|_k$, $\hat{\theta}|_k$ is a MMSE estimate. $\triangleleft$

It follows from 3. and 2. that, if $\binom{\Theta}{z^{(k)}}$ is jointly gaussian then (25) is a necessary and sufficient condition for a centered $\tilde{\Theta}|_k$ to be a MMSE estimate.

If $\binom{\Theta}{z^{(k)}}$ is not jointly gaussian but $\Psi_{z^{(k)}}$ is nonsingular (for example, when $z^{(k)}$ is gaussian), then (25) is a necessary and sufficient condition for a $\tilde{\Theta}|_k \in \{\Theta|_k\}$ (i.e. affine with respect to $z^{(k)}$ and centered) to be a MMSE estimate.

KALMAN FILTER

Consider

$$\boxed{\begin{aligned} x(j+1) &= A(j)\,x(j) + B(j)\,u(j) \qquad j = 0,\dots,k, \\ z(j) &= C(j)\,x(j) + v(j) \qquad j = 1,\dots,k, \end{aligned}}$$

(27)

where $x(j) \in \mathbb{R}^n$, $u(j) \in \mathbb{R}^p$, $z(j) \in \mathbb{R}^q$, the sequences $\{u(j)\}$ and $\{v(j)\}$ are white, gaussian, mutually uncorrelated, $E[v(j)] = 0$, $E[u(j)] = 0$ $\forall j$, known $\Psi_{v(j)}$ and $\Psi_{u(j)}$, $\forall j$. Moreover, $x(0)$ is gaussian with mean $\bar{x}(0)$ and covariance $\Psi_{x(0)}$ (both known), uncorrelated with each $\{v(j)\}$ and $\{u(j)\}$. We refer to these assumptions as standard Kalman assumptions.

The MMSE estimate $x(i|k)$ of $x(i)$ is, as known,

$$\tilde{x}(i|k) = E[x(i) \mid z^{(k)}]$$

Since, by the standard Kalman assumptions on (27), $p\left(\frac{\theta}{z^{(k)}}\right)$ is gaussian (apply 2.1 and 2.2), then it follows from 2. pg 65 and (25) that

$$\begin{cases} E[x(i) \mid z^{(k)}] = \overline{x}(i) + \Psi_{x(i)z^{(k)}} \Psi^{-1}_{z^{(k)}} \left(z^{(k)} - \overline{z}^{(k)}\right) \\ \Psi_{x(i)\mid z^{(k)}} = \Psi_{x(i)} - \Psi_{x(i)z^{(k)}} \Psi^{-1}_{z^{(k)}} \Psi_{z^{(k)}x(i)} \end{cases} \quad (28)$$

where the $\overline{\cdot}$ denotes expectation.

It is clear that, for increasing i or k, computational problems arise and a recursive implementation of $E[x(i) \mid z^{(k)}]$ is much preferred to (28).

To obtain a recursive implementation of (28) we have to guarantee for the recursive filter three main properties:

1 affinity with respect to $z^{(k)}$

2 centering

3 optimality (in the sense of MMSE)

We focus on the case $i=k$ (FILTERING).

Our recursive structure is

$$\boxed{\tilde{x}(j+1|j+1) = G(j+1)\tilde{x}(j|j) + F(j+1)z(j+1), \quad j=0,.., k-1 \quad (29)}$$

which clearly guarantees 1 (affinity with respect to $z^{(k)}$). Next, we want to guarantee 2 (centering). ~~Assume~~ that $\tilde{x}(j|j)$ is centered. Let's see how $G(j+1)$ should be chosen so that $\tilde{x}(j+1|j+1)$ is centered.

We have, using (29) and (27),

$$E\left[\tilde{x}(j+1|j+1)\right] =$$

$$= E\left[G(j+1)\,\tilde{x}(j|j) + F(j+1)\,z(j+1)\right] =$$

$$= G(j+1)\,E\left[x(j)\right] + F(j+1)\,E\left[C(j+1)\,x(j+1) + v(j+1)\right]$$

use centering of $\tilde{x}(j|j)$, i.e. $E[x(j)] = E[\tilde{x}(j|j)]$

$$= G(j+1)\,E[x(j)] + F(j+1)\,C(j+1)\,A(j)\,E[x(j)]$$

$$+ F(j+1)\left[C(j+1)\,B(j)\,E[m(j)] + E[v(j+1)]\right]$$

$$= \left[G(j+1) + F(j+1)\,C(j+1)\,A(j) - A(j) + A(j)\right]E[x(j)]$$

$$= A(j)\,E[x(j)] = E\left[A(j)\,x(j) + B(j)\,m(j)\right] = E\left[x(j+1)\right]$$

by taking $\boxed{G(j+1) := A(j) - F(j+1)\,C(j+1)\,A(j)}$ (30)

which proves that $\tilde{x}(j+1|j+1)$ is centered.

Finally, we have to guarantee that also $\tilde{x}(0|0)$ is centered. This automatically requires that

$$E[\tilde{x}(0|0)] := E[x(0)] = \bar{x}(0) \quad (31)$$

For example, the filter (29) should be initialised as follows

$$\boxed{\tilde{x}(0|0) = E[x(0)] \qquad (32)}$$

which guarantees (31). This also gives the initial value of $\Psi_{\tilde{e}(0|0)}$

$$\begin{aligned}
\Psi_{\tilde{e}(0|0)} &= E\left[(x(0)-\tilde{x}(0|0))(x(0)-\tilde{x}(0|0))^T\right] \\
&= E\left(x(0)-\bar{x}(0))(x(0)-\bar{x}(0))^T\right] \\
&= \Psi_{x(0)} \qquad (33)
\end{aligned}$$

The last property to be guaranteed for (29) is ③ optimality (in the sense of MMSE. This is guaranteed through the necessary and sufficient condition (25)

$$E[\tilde{e}(j|j)\, z^{(j)T}] = 0, \; j = 1, \ldots, k,$$

From now on we will assume that the choice of $F(j+1)$ will be such that

$$\boxed{\det\left(I - F^T(j+1)\,C^T(j+1)\right) \neq 0, \quad j=0,..,k-1. \quad (34)}$$

Assume that

$$E\left[\tilde{e}(j|j)\, z^{(j)^T}\right] = 0 \qquad (35)$$

for some $j = 1, .., k-1$, and prove

$$E\left[\tilde{e}(j+1|j+1)\, z^{(j+1)^T}\right] = 0 \qquad (36)$$

First, prove

$$E\left[\tilde{e}(j+1|j+1)\, z^T(\tau)\right] = 0, \quad \tau = 1, .., j \qquad (37)$$

As a matter of fact, using (29), (27) and (30)

$$E\left[\tilde{e}(j+1|j+1)\, z^T(\tau)\right] =$$

$$= E\Big[\Big(A(j)\,x(j) + B(j)\,m(j) - \big(A(j) - F(j+1)\,C(j+1)\,A(j)\big)\cdot$$

$$\tilde{x}(j|j) - F(j+1)\,z(j+1)\Big)\, z^T(\tau)\Big] =$$

$$= E\{[(A(j) - F(j+1)C(j+1)A(j))\tilde{e}(j|j)$$
$$+ (B(j) - F(j+1)C(j+1)B(j))m(j)$$
$$- F(j+1)v(j+1)]z^T(\tau)\} = 0$$

since $E[\tilde{e}(j|j)z^T(\tau)] = 0$, $\tau = 1, .., j$ (by (35)) and since $m(j)$ and $v(j+1)$ are uncorrelated with $z(\tau)$, $\tau = 1, .., j$, since this is a linear combination of $x(0), m(0), \ldots, m(j-1), v(1), .., v(j)$.

This proves (37). Next, prove

$$E[\tilde{e}(j+1|j+1)z^T(j+1)] = 0 \quad (38)$$

Equivalentely, on account of (34), we prove

$$0 = E[\tilde{e}(j+1|j+1)z^T(j+1)(I - F^T(j+1)C^T(j+1))]$$
$$(39)$$

We have, using (29) and (27)

$$E\left[\tilde{e}(j+1|j+1)z^T(j+1)\left(I - F^T(j+1)C^T(j+1)\right)\right]$$

$$= E\left\{\tilde{e}(j+1|j+1)\left[z^T(j+1) - \tilde{x}^T(j|j)G^T(j+1)C^T(j+1) - z^T(j+1)F^T(j+1)C^T(j+1)\right]\right\}$$

N.B. use $E\left[\tilde{e}(j+1|j+1)\tilde{x}^T(j|j)\right] = 0$ by (37) since $\tilde{x}(j|j)$ is a linear combination of $z(1), \dots, z(j)$

$$= E\left\{\tilde{e}(j+1|j+1)\left[z(j+1) - C(j+1)\tilde{x}(j+1|j+1)\right]^T\right\}$$

$$= E\left\{\tilde{e}(j+1|j+1)\tilde{e}^T(j+1|j+1)C^T(j+1)\right\}$$

$$+ E\left\{\tilde{e}(j+1|j+1)\nu^T(j+1)\right\}$$

$$= \Psi_{\tilde{e}(j+1|j+1)}C^T(j+1) - E\left[\tilde{x}(j+1|j+1)\nu^T(j+1)\right]$$

N.B. since $E\left[x(j+1)\nu^T(j+1)\right] = 0$ being $x(j+1)$ a linear combination of $x(0), m(0), \dots, m(j)$

$$= \Psi_{\tilde{e}(j+1|j+1)}C^T(j+1) - E\left[F(j+1)z(j+1)\nu^T(j+1)\right]$$

N.B. since $E\left[\tilde{x}(j|j)\nu^T(j+1)\right] = 0$ being $\tilde{x}(j|j)$ a linear combination of $x(0), m(0), \dots, m(j-1), \nu(1), \dots, \nu(j)$.

$$= \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) - F(j+1)\Psi_{\nu(j+1)}$$

N.B

use $E[x(j+1)\nu^T(j+1)] = 0$ since $x(j+1)$ is a linear combination of $x(0), m(0), \ldots, m(j)$

Therefore, by choosing

$$\boxed{F(j+1) = \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) \Psi^{-1}_{\nu(j+1)}} \quad (40)$$

We have only to prove that

$$E[\tilde{e}(1|1) z^T(1)] = 0$$

This can be shown exactly as (38).
We have also to prove that the choice (40) is compatible with (34). First, we derive the recursive expression for

$\Psi_{\hat{e}(j+1|j+1)}$.

One has from (29) and (27)

$$\Psi_{\tilde{e}(j+1|j+1)} = E\left\{\tilde{e}(j+1|j+1)\tilde{e}^T(j+1|j+1)\right\}$$

$$= E\Big\{\big[A(j)x(j)+B(j)m(j)-A(j)\tilde{x}(j|j)$$
$$-F(j+1)z(j+1)+F(j+1)C(j+1)A(j)\tilde{x}(j|j)\big]\cdot$$
$$\cdot\big[A(j)x(j)+B(j)m(j)-A(j)\tilde{x}(j|j)$$
$$-F(j+1)z(j+1)+F(j+1)C(j+1)A(j)\tilde{x}(j|j)\big]^T\Big\}$$

$$= E\Big\{\big[(I-F(j+1)C(j+1))A(j)\tilde{e}(j|j)$$
$$+(I-F(j+1)C(j+1))B(j)m(j)-F(j+1)\nu(j+1)\big]\cdot$$
$$\cdot\big[(I-F(j+1)C(j+1))A(j)\tilde{e}(j|j)$$
$$+(I-F(j+1)C(j+1))B(j)m(j)-F(j+1)\nu(j+1)\big]^T\Big\}$$

$$= (I-F(j+1)C(j+1))\big[\underbrace{A(j)\Psi_{\tilde{e}(j|j)}A^T(j)+B(j)\Psi_{u(j)}B^T(j)}_{:=\Gamma(\Psi_{\tilde{e}(j|j)})}\big]\cdot$$
$$\cdot(I-F(j+1)C(j+1))^T + F(j+1)\Psi_{\nu(j+1)}F^T(j+1)$$

Using (40)

$$\Psi_{\tilde{e}(j+1|j+1)} = \left(I-\Psi_{\tilde{e}(j+1|j+1)}C^T(j+1)\Psi^{-1}_{\nu(j+1)}C(j+1)\right)\cdot$$
$$\cdot\Gamma(\Psi_{\tilde{e}(j|j)})\left(I-\Psi_{\tilde{e}(j+1|j+1)}C^T(j+1)\Psi^{-1}_{\nu(j+1)}C(j+1)\right)^T$$

$$+ \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) \Psi^{-1}_{v(j+1)} \Psi_{\tilde{e}(j+1|j+1)}$$

which gives

$$\Psi_{\tilde{e}(j+1|j+1)} \left(I - C^T(j+1) \Psi^{-1}_{v(j+1)} C(j+1) \Psi_{\tilde{e}(j+1|j+1)} \right)$$

$$= \left(I - \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) \Psi^{-1}_{v(j+1)} C(j+1) \right) \Gamma(\Psi_{\tilde{e}(j|j)}) \cdot$$

$$\cdot \left(I - C^T(j+1) \Psi^{-1}_{v(j+1)} C(j+1) \Psi_{\tilde{e}(j+1|j+1)} \right) \quad (41)$$

If we assume that

$$\boxed{\det \left(I - C^T(j+1) \Psi^{-1}_{v(j+1)} C(j+1) \Psi_{\tilde{e}(j+1|j+1)} \right) \neq 0 \quad (42)}$$

then it follows from (41)

$$\Psi_{\tilde{e}(j+1|j+1)} = \left(I - \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) \Psi^{-1}_{v(j+1)} C(j+1) \right) \cdot$$

$$\cdot \Gamma(\Psi_{\tilde{e}(j|j)})$$

and

$$\Psi_{\tilde{e}(j+1|j+1)}\Big(I + C^T(j+1)\,\Psi^{-1}_{v(j+1)}\,C(j+1)\,\cdot$$
$$\cdot\,\Gamma(\Psi_{\tilde{e}(j|j)})\Big) = \Gamma(\Psi_{\tilde{e}(j|j)})$$

Assuming also that

$$\boxed{0 \neq \det\left(I + C^T(j+1)\,\Psi^{-1}_{v(j+1)}\,C(j+1)\,\Gamma(\Psi_{\tilde{e}(j|j)})\right) \qquad (43)}$$

we conclude with

$$\boxed{\begin{aligned}\Psi_{\tilde{e}(j+1|j+1)} &= \Gamma(\Psi_{\tilde{e}(j|j)})\,\cdot \\ &\cdot\left(I + C^T(j+1)\,\Psi^{-1}_{v(j+1)}\,C(j+1)\cdot\Gamma(\Psi_{\tilde{e}(j|j)})\right)^{-1} \qquad (44)\end{aligned}}$$

together with $\Psi_{\tilde{e}(0|0)} = \Psi_{x(0)}$ (see (33)).
We only have to prove (34), (42) and (43).

First, we prove (43).

The matrix

$$I + C^T(j+1)\Psi^{-1}_{v(j+1)}C(j+1)\,\Gamma(\Psi_{\breve{e}(j|j)})$$

has the form $I + PQ$, where

$P := C^T(j+1)\Psi^{-1}_{v(j+1)}C(j+1)$ and $Q := \Gamma(\Psi_{\breve{e}(j|j)})$

are positive semidefinite and symmetric.

In order to prove (43) we prove that $I + PQ$ is nonsingular. Assume that there exists $\bar{x} \neq 0$ such that

$$(I + PQ)\,\bar{x} = 0 \qquad (45)$$

From this $\bar{x}^T Q(I + PQ)\bar{x} = 0$ and then

$$\begin{cases} \bar{x}^T Q \bar{x} = 0 \\ \bar{x}^T Q P Q \bar{x} = 0 \end{cases} \qquad (46)$$

since Q and P are positive semidefinite.

The first of (43) implies $Q\bar{x} = 0$. Indeed, Q is symmetric and positive semidefinite.

There exists orthogonal T (i.e. $T^T = T^{-1}$) which diagonalizes Q:

$$TQT^{-1} = \Lambda = \mathrm{diag}\{\lambda_1 \dots \lambda_n\}$$

and $\lambda_i \geq 0 \ \forall i$. Therefore from (46)

$$0 = \bar{x}^T Q \bar{x} = \bar{x}^T T^T \Lambda T \bar{x} = \underbrace{\bar{x}^T T^T}_{z^T} \Lambda \underbrace{T\bar{x}}_{z}$$

$$= z^T \Lambda z$$

which implies $0 = \Lambda z$ and then

$$0 = T^{-1} \Lambda z = T^{-1} \Lambda T \bar{x} = Q\bar{x}$$

which proves what claimed. Now, using $0 = Q\bar{x}$ in (45) we obtain $\bar{x} = 0$. This prove the nonsingularity of $I + PQ$ and therefore (43).

Next, we prove (42). As a consequence of (43) and (44)

$$I - C^T(j+1)\Psi_{v(j+1)}^{-1} C(j+1)\Psi_{\tilde{e}(j+1|j+1)}$$

$$= I - C^T(j+1)\Psi_{v(j+1)}^{-1} C(j+1)\Gamma(\Psi_{\tilde{e}(j|j)}) \cdot$$

$$\cdot \left(I + C^T(j+1)\Psi_{v(j+1)}^{-1} C(j+1)\Gamma(\Psi_{\tilde{e}(j|j)})\right)^{-1}$$

$$= \Big[I + C^T(j+1)\Psi_{v(j+1)}^{-1} C(j+1)\Gamma(\Psi_{\tilde{e}(j|j)})$$

$$- C^T(j+1)\Psi_{v(j+1)}^{-1} C(j+1)\Gamma(\Psi_{\tilde{e}(j|j)}) \Big] \cdot$$

$$\cdot \left(I + C^T(j+1)\Psi_{v(j+1)}^{-1} C(j+1)\Gamma(\Psi_{\tilde{e}(j|j)})\right)^{-1}$$

$$= \left(I + C^T(j+1)\Psi_{v(j+1)}^{-1} C(j+1)\Gamma(\Psi_{\tilde{e}(j|j)})\right)^{-1}$$

where the last matrix is nonsingular by (43). Therefore, also (42) is true.

Finally, we prove (34). Recalling (40)

$$\det(I - F^T(j+1)C^T(j+1))$$

$$= \det(I - C^T(j+1)F^T(j+1))$$

N.B. This follows from $\det(I_{p\times p} + RS) = \det(I_{n\times n} + SR)$ where $R \in \mathbb{R}^{p\times n}$, $S \in \mathbb{R}^{n\times p}$.

97

$$= \det\left(I - C^T(j+1)\Psi^{-1}_{\nu(j+1)}C(j+1)\Psi_{\tilde{e}(j+1|j+1)}\right)$$

$$\neq 0$$

$\uparrow$

as a consequence of (43) and (44) $\triangleleft$

In conclusion, the equation for the kalman filter are given by

$$\begin{cases} \tilde{x}(j+1|j+1) = A(j)\tilde{x}(j|j) + \\ + \underbrace{\Psi_{\tilde{e}(j+1|j+1)}C^T(j+1)\Psi^{-1}_{\nu(j+1)}}_{:= F(j+1)}\left[z(j+1) - C(j+1)A(j)\tilde{x}(j|j)\right] \quad j = 0, \dots, k-1, \\ \tilde{x}(0|0) = \bar{x}(0) \end{cases}$$

$$\begin{cases} \Psi_{\tilde{e}(j+1|j+1)} = \Gamma(\Psi_{\tilde{e}(j|j)})\left[I + C^T(j+1)\Psi^{-1}_{\nu(j+1)}C(j+1)\Gamma(\Psi_{\tilde{e}(j|j)})\right]^{-1} \\ \qquad j = 0, \dots, k-1, \\ \Gamma(\Psi_{\tilde{e}(j|j)}) = A(j)\Psi_{\tilde{e}(j|j)}A^T(j) + B(j)\Psi_{u(j)}B^T(j) \\ \Psi_{\tilde{e}(0|0)} = \Psi_{x(0)} \end{cases} \quad (47)$$

$F(j+1)$ is also called Kalman gain and

$$e(j+1) := z(j+1) - C(j+1)A(j)\tilde{x}(j|j)$$

is the innovation sequence of the Kalman filter.

It should be remembered that if gaussianity is not assumed for $x(0)$, $\{m(j)\}$ and $\{v(j)\}$, the Kalman filter gives the optimal MMSE estimate among all the possible affine and centered estimates (by 3. pg. 78).

Moreover, the innovation sequence is gaussian and white, with zero mean.

Indeed,

$$E[e(j)] =$$

$$= E[z(j) - C(j)A(j-1)\tilde{x}(j-1|j-1)] =$$

$$= E[z(j)] - C(j)A(j-1)E[\tilde{x}(j-1|j-1)] =$$

$$= E[z(j)] - C(j)A(j-1)E[x(j-1)]$$

$$= E[z(j) - C(j)(A(j-1)x(j-1) + B(j-1)m(j-1))]$$

$$= E[z(j) - C(j)x(j) - v(j)] =$$

$$= E[C(j)\tilde{e}(j|j)] = 0 \Rightarrow \text{zero mean.}$$

Moreover, for each pair (i,j) (with $i<j$)

$$E[e(i)e^T(j)] =$$

$$= E[(z(i) - C(i)A(i-1)\tilde{x}(i-1|i-1))(z(j) - C(j)A(j-1)\cdot$$

$$\cdot\tilde{x}(j-1|j-1))^T] =$$

$$= E[(z(i) - C(i)A(i-1)\tilde{x}(i-1|i-1))\cdot(C(j)A(j-1)\tilde{e}(j-1|j-1)$$

$$+ C(j)B(j-1)m(j-1) + v(j))^T]$$

$$= 0$$

N.B.
since $E[\tilde{e}(j-1|j-1)z(i)^T] = 0$, $i \leq j-1$,
by optimality condition (25) and since
$E[\tilde{x}(i-1|i-1)\tilde{e}^T(j-1|j-1)] = 0$ being $\tilde{x}(i-1|i-1)$
a linear combination of $z(1), \dots, z(i-1)$, with
$i-1 \leq j-1$.

Similarly for $i=j$

$$E[e(i)e^T(i)] =$$

$$= C(i)A(i-1)\Psi_{\tilde{e}(i-1|i-1)}A^T(i-1)C^T(i)$$

$$+ C(i)B(i-1)\Psi_{u(i-1)}B^T(i-1)C^T(i) .$$

Finally, $\{e(i)\}$ is also gaussian because is a linear combination of $x(0)$, $\{m(i)\}$ and $\{v(i)\}$ which are white, uncorrelated and gaussian.

In conclusion, we wan to prove one key property of $\{e(i)\}$, i.e. $e(j+1)$ is uncorrelated with $z(1), .., z(j)$:

$$\boxed{E[e(j+1)z^T(r)] = 0 , \quad r = 1,..,j \qquad (48)}$$

Indeed,

$$E\left[e(j+1)\,z^T(\tau)\right] = E\left[\left(z(j+1) - C(j+1)A(j)\cdot\right.\right.$$

$$\left.\left.\cdot\,\tilde{x}(j|j)\right) z^T(\tau)\right] =$$

$$= E\left[\left(C(j+1)x(j+1) + v(j+1) - C(j+1)A(j)\tilde{x}(j|j)\right)\cdot\right.$$

$$\left.\cdot\, z^T(\tau)\right] =$$

$$= E\left[\left(C(j+1)A(j)\tilde{e}(j|j) + C(j+1)B(j)m(j)\right.\right.$$

$$\left.\left.+\,v(j+1)\right)\cdot z^T(\tau)\right] = 0$$

since $z(\tau)$, $\tau = 1, .., j$, is uncorrelated with $\tilde{e}(j|j)$ (by optimality (25)), $m(j)$ (being a linear combination of $x(0)$, $m(0), .., m(j-1), v(1), .., v(j)$) and $v(j+1)$ (for the above reason).

PREDITTORE DI KALMAN

With reference to (27), we want to obtain the MMSE estimate $\tilde{\theta}(i|k)$ in the case $k < i$ (PREDICTION).

First of all, by (27) we obtain

$$x(i) = \phi(i,k)\,x(k) + \sum_{j=k}^{i-1} \phi(i,j+1)\,B(j)\,m(j)$$

where $\phi(i,j) = \begin{cases} A(i-1)\cdots A(j) & i > j \\ I & i = j \end{cases}$.

We want to prove that

$$\tilde{x}(i|k) = \phi(i,k)\tilde{x}(k|k) \quad , \quad i > k$$

$$\Psi_{\tilde{e}(i|k)} = \phi(i,k)\Psi_{\tilde{e}(k|k)}\phi^T(i,k)$$

$$+ \sum_{j=k}^{i-1} \phi(i,j+1)B(j)\Psi_{u(j)}B^T(j)\phi^T(i,j+1),$$

where $\tilde{x}(k|k)$ is the MMSE estimate of $x(k)$ (49)

is the MMSE prediction of $x(i)$ (with k measurements).

First, we prove that $\tilde{x}(i|k)$ is centered.

$$E[\tilde{x}(i|k)] = E[\phi(i,k)\tilde{x}(k|k)] =$$

$$= \phi(i,k)E[\tilde{x}(k|k)] = \phi(i,k)E[x(k)]$$

since $\tilde{x}(k|k)$ is centered

$$= E\left[\phi(i,k)x(k) + \sum_{j=k}^{i-1}\phi(i,j+1)B(j)m(j)\right]$$

$$= E[x(i)]$$

which proves that $\tilde{x}(i|k)$ is centered. Next, we prove the optimality of $\tilde{x}(i|k)$ by using condition (25):

$$E\left[\tilde{e}(i|k)\, z^{(k)^T}\right] = 0$$

where $\tilde{e}(i|k) := x(i) - \tilde{x}(i|k)$. Indeed,

$$E\left[\tilde{e}(i|k)\, z^T(\tau)\right] = \qquad , \tau = 1, \dots, k$$

$$= E\left[\left(x(i) - \tilde{x}(i|k)\right) z^T(\tau)\right] =$$

$$= E\left[\left(\phi(i,k)\,\tilde{e}(k|k) + \sum_{j=k}^{i-1} \phi(i,j+1) B(j) m(j)\right) \cdot z^T(\tau)\right] = \phi(i|k)\, E\left[\tilde{e}(k|k)\, z^T(\tau)\right]$$

$$+ \sum_{j=k}^{i-1} \phi(i,j+1) B(j)\, E\left[m(j)\, z^T(\tau)\right]$$

$$= 0$$

since $E\left[\tilde{e}(k|k)\, z^{(k)^T}\right] = 0$ by optimality of $\check{x}(k|k)$ and since $z(\tau)$, $\tau = 1, \dots, k$, are linear combination of $x(0)$, $m(0), \dots, m(k-1)$,

$v(1), \ldots, v(k)$ which are all uncorrelated from $m(k+1), \ldots, m(i-1)$. $\triangleleft$

Note that the MMSE one-step prediction (i.e. $i = k+1$) is given by

$$\boxed{\tilde{x}(k+1|k)} = \phi(k+1,k)\tilde{x}(k|k) = \boxed{A(k)\tilde{x}(k|k)} \qquad (50)$$

$$\begin{aligned}\boxed{\Psi_{\tilde{e}(k+1|k)}} &= \phi(k+1,k)\Psi_{\tilde{e}(k|k)}\phi^T(k+1,k) \\ &+ \phi(k+1,k+1)B(k)\Psi_{u(k)}B^T(k) \cdot \\ &\cdot \phi^T(k+1,k+1) \\ &= A(k)\Psi_{\tilde{e}(k|k)}A^T(k) + B(k)\Psi_{u(k)}B^T(k) \\ &= \boxed{\Gamma(\Psi_{\tilde{e}(k|k)})} \qquad (51)\end{aligned}$$

In what follows,
we study some variants of the
Kalman filter.

(A) $E[m(j)] \neq 0$, $E[v(j)] \neq 0$
for some $j \geq 0$, and
$\{m(j)\}$ has a deterministic
component, i.e. $m(j) = m_s(j) + m_d(j)$
where $\{m_s(j)\}$ is random, and
$\{m_d(j)\}$ deterministic and known.

Here, we define

$$m_s'(j) = m_s(j) - \bar{m}_s(j)$$

$$v'(j) = v(j) - \bar{v}(j)$$

$$x'(0) = x(0) - \bar{x}(0)$$

$$m_d'(j) = \bar{m}_s(j) + m_d(j)$$

Equation (27) are re-written as

① $x_d(j+1) = A(j)x_d(j) + B(j)m'_d(j), \quad j=0,..,k-1,$

② $\begin{cases} x_s(j+1) = A(j)\,x_s(j) + B(j)\,m'_s(j), & j=0,..,k-1, \\ z_s(j) = C(j)\,x_s(j) + v'(j) & j=1,..,k \\ \quad (= z(j) - C(j)x_d(j) - \bar{v}(j)) & \end{cases}$

③ $x(j) = x_s(j) + x_d(j), \qquad j=1,..,k$

where $E[m'_s(j)] = 0, \; E[v'(j)] = 0 \;\; \forall j,$

$E[x(0)] = x_d(0) + E[x_s(0)]$

$= x_d(0) + \bar{x}_s(0)$.

The kalman filter is applied to ② and a MMSE estimate $\tilde{x}_s(j|j)$ is obtained. The estimate $\tilde{x}(j|j) = x_d(j) + \tilde{x}_s(j|j)$ is a MMSE estimate of $x(j)$. The kalman filter for ② is implemen ted using the measurements

$$z_s(j) = z(j) - C(j)x_d(j) - \bar{v}(j)$$

and

$$\bar{x}_s(0) = \bar{x}(0) - x_d(0)$$

$\boxed{\text{B} \quad \{v(j)\} \text{ and } \{m(j)\} \text{ are mutually correlated}}$

For this problem we consider in place of (27) the equations

$$x(j+1) = A(j)x(j) + B(j)\,m(j) + D(j)v(j)$$
$$j = 0, \ldots, k-1$$

$$z(j) = C(j)x(j) + v(j) \quad j = 0, \ldots, k \qquad (52)$$

where the input sequence has a common component with the measurement noise sequence $\{v(j)\}$

We assume all the standard kalman assumptions for $x(0)$, $\{v(j)\}$ and $\{m(j)\}$.

Here, instead of (29) we consider the following recursive structure

$$\tilde{x}(j+1|j+1) = G(j+1)\tilde{x}(j+1|j) + F(j+1)z(j+1) \qquad (53)$$

which again guarantees (1) (affinity of $\tilde{x}(j+1|j+1)$ with respect to $z^{(j+1)}$).
We have to require centering of $\tilde{x}(j+1|j+1)$ and its optimality (via (25)).
Centering requirement leads to

$$\boxed{G(j+1) = I - F(j+1)C(j+1)} \qquad (54)$$

(instead of (30) in the case of uncorrelation between $\{m(j)\}$ and $\{v(j)\}$)

Optimality requirement leads to

$$\boxed{F(j+1) = \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1)\, \Psi^{-1}_{v(j+1)} \quad (55)}$$

(as in (40) for the case of uncorrelation between $\{m(j)\}$ and $\{v(j)\}$).

We have to evaluate $\tilde{x}(j+1|j)$:

$$\tilde{x}(j+1|j) = E[x(j+1) \mid z^{(j)}] =$$

$$= A(j)E[x(j) \mid z^{(j)}] + B(j)E[m(j) \mid z^{(j)}]$$

$$+ D(j)E[v(j) \mid z^{(j)}] =$$

$$= A(j)\tilde{x}(j|j) + D(j)E[(z(j) - C(j)x(j)) \mid z^{(j)}]$$

N.B.

$E[m(j) \mid z^{(j)}] = E[m(j)] = 0$ since $z^{(j)}$ is a linear combinations of $x(0), m(0), \ldots, m(j-1)$ $v(1), \ldots, v(j)$, all gaussian and uncorrelated from $m(j)$, therefore $m(j)$ and $z^{(j)}$ are (gaussian and) independent.

$$= A(j)\tilde{x}(j|j) + D(j)\left[z(j) - C(j)\tilde{x}(j|j)\right] \quad (56)$$

N.B

$E[z(j)|z^{(j)}] = z(j)$ and

$E[x(j)|z^{(j)}] = \tilde{x}(j|j)$

Moreover, using (53) and (54)

$$\tilde{e}(j+1|j+1) = x(j+1) - \tilde{x}(j+1|j+1)$$

$$= [I - F(j+1)C(j+1)]\tilde{e}(j+1|j)$$

$$- F(j+1)\,v(j+1)$$

and, consequentely,

$$\Psi_{\tilde{e}(j+1|j+1)} = (I - F(j+1)C(j+1))\cdot\Psi_{\tilde{e}(j+1|j)}\cdot$$

$$\cdot(I - F(j+1)C(j+1))^T + F(j+1)\Psi_{v(j+1)}F^T(j+1)$$

By direct calculation, using (56) and (52)

$$\Psi_{\tilde{e}(j+1|j)} = E\Big[\big(x(j+1)-\tilde{x}(j+1|j)\big)\cdot$$

$$\cdot\big(x(j+1)-\tilde{x}(j+1|j)\big)^T\Big]$$

$$= E\Big\{\Big(\big[A(j)-D(j)C(j)\big]\tilde{e}(j|j)+B(j)m(j)\Big)\cdot$$

$$\Big(\big[A(j)-D(j)C(j)\big]\tilde{e}(j|j)+B(j)m(j)\Big)^T\Big]$$

$$= \big[A(j)-D(j)C(j)\big]\Psi_{\tilde{e}(j|j)}\big[A(j)-D(j)C(j)\big]^T$$

$$+ B(j)\Psi_{u(j)}B^T(j)\,. \qquad (57)$$

Collecting (53), (54), (56) and (57) we arrive at

$$\boxed{\begin{aligned}&\tilde{x}(j+1|j+1) = A(j)\tilde{x}(j|j) + F(j+1)\big[z(j+1)\\ &- C(j+1)A(j)\tilde{x}(j|j)\big] +\\ &+\big[I - F(j+1)C(j+1)\big]D(j)\big[z(j)-C(j)\tilde{x}(j|j)\big]\\ &\qquad\qquad j=0,\dots,k-1,\\ &\tilde{x}(0|0)=\bar{x}(0) \qquad\qquad (58)\end{aligned}}$$

$$F(j+1) = \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) \Psi^{-1}_{v(j+1)} \tag{59}$$

$$\begin{aligned}\Psi_{\tilde{e}(j+1|j+1)} = {} & [I - F(j+1)C(j+1)] \cdot \Big\{ \\ & [A(j) - D(j)C(j)] \Psi_{\tilde{e}(j|j)} [A(j) - D(j)C(j)]^T \\ & + B(j) \Psi_{m(j)} B^T(j) \Big\} \cdot [I - F(j+1)C(j+1)]^T \\ & + F(j+1) \Psi_{v(j+1)} F^T(j+1)\end{aligned} \tag{60}$$

$$\Psi_{\tilde{e}(0|0)} = \Psi_{x(0)}$$

EXTENDED KALMAN FILTER

Here we consider nonlinear models

$$\boxed{\begin{aligned} x(j+1) &= f(x(j), j) + B(j)w(j) \qquad j = 0, \ldots, k-1 \\ z(j) &= g(x(j), j) + v(j) \qquad j = 1, \ldots, k \end{aligned} \qquad (61)}$$

and we retain for $x(0)$, $\{w(j)\}$ and $\{v(j)\}$ the standard Kalman assumptions. We want to calculate some approximation of the MMSE estimate $E[x(k) \mid z^{(k)}]$.

The strategy is to calculate (on-line or off-line) a reference sequence $\{x^*(j)\}$ ...!

At each time j, we consider the Taylor expansion of f and g around $x^*(j)$ (up to the first order):

$$\begin{cases} x(j+1) \simeq A^*(j)x(j) + B(j)u(j) + m^*(j) \\ z(j) \simeq C^*(j)x(j) + v(j) + d^*(j) \end{cases} \quad (62)$$

where $A^*(j) := \left.\frac{\partial f(x,j)}{\partial x}\right|_{x=x^*(j)}$

$$C^*(j) := \left.\frac{\partial g(x,j)}{\partial x}\right|_{x=x^*(j)}$$

$$m^*(j) := f(x^*(j),j) - A^*(j)x^*(j)$$

$$d^*(j) := g(x^*(j),j) - C^*(j)x^*(j)$$

A Kalman filter for (62) can be implemented as pointed out in (A) pg. 106, ~~if~~ $\{m^*(j)\}$ and $\{d^*(j)\}$ are deterministic sequences, for example, when $\{x^*(j)\}$ is deterministic.

A deterministic generation of $\{x^*(j)\}$ can be implemented as follows:

$\{x^*(j)\}$ is such that $\forall j \geq 0$

$$\begin{cases} x^*(j+1) = f(x^*(j), j) \\ x^*(0) = \bar{x}(0) \end{cases}$$

or, in other words, $x^*(j)$ is the state trajectory ensuing from $\bar{x}(0)$.

If we separate the deterministic and random component of the state

$$x(j) = x_s(j) + x_d(j)$$

then (62) can be seen as the superposition of

(197)

$$
\begin{cases}
x_d(j+1) = A^*(j)\, x_d(j) + m^*(j), \\
\qquad x_d(0) = \bar{x}(0), \\
\qquad j = 0, .., k-1 \\
z_d(j) = C^*(j)\, x_d(j) + d^*(j), \quad j = 1, .., k,
\end{cases} \tag{62}
$$

and

$$
\begin{cases}
x_s(j+1) = A^* x_s(j) + B(j)\, m(j), \quad j = 0, .., k-1, \\
\qquad x_s(0) = x(0) - \bar{x}(0), \\
z_s(j) = C^*(j)\, x_s(j) + v(j), \quad j = 1, .., k.
\end{cases} \tag{63}
$$

The kalman filter for (63) gives

$$
\begin{cases}
\tilde{x}_s(j+1|j+1) = A^* \tilde{x}_s(j|j) + \Psi^*_{\tilde{e}(j+1|j+1)}\, C^{*T}(j+1) \cdot \\
\qquad \cdot \Psi^{-1}_{v(j+1)} \left[z_s(j+1) - C^*(j+1) A^*(j) \tilde{x}_s(j|j) \right] \\
\tilde{x}_s(0|0) = \bar{x}_s(0) = 0 \qquad , j = 0, .., k-1 \\
\Psi_{\tilde{e}(j+1|j+1)} = \Gamma(\Psi_{\tilde{e}(j|j)}) \left[I + C^{*T}(j+1) \Psi^{-1}_{v(j+1)} C^*(j+1) \Gamma(\Psi_{\tilde{e}(j|j)}) \right]^{-1} \\
\Psi_{\tilde{e}(0|0)} = \Psi_{x(0)} \qquad j = 0, .., k-1.
\end{cases}
$$

Finally we take as estimate of $x(j)$

$$\boxed{\tilde{x}(j+1|j+1)} = x_d(j+1) + \tilde{x}_s(j+1|j+1)$$

$$= x_d(j+1) + A^*(j)\tilde{x}_s(j|j) + \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) \cdot$$
$$\cdot \Psi^{-1}_{v(j+1)} [z_s(j+1) - C^*(j+1)A^*(j)\tilde{x}_s(j|j)]$$

$$= A^*(j)\tilde{x}(j|j) + m^*(j) + \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) \cdot$$
$$\cdot \Psi^{-1}_{v(j+1)} \{z(j+1) - C^*(j+1)[A^*(j)\tilde{x}(j|j) + m^*(j)] + d^*(j+1)\}$$

$$\boxed{\simeq \tilde{x}(j+1|j) + \Psi_{\tilde{e}(j+1|j+1)} C^T(j+1) \cdot \Psi^{-1}_{v(j+1)} \{z(j+1) - g[\tilde{x}(j+1|j), j+1)]\}}\ ,$$

where $\boxed{\tilde{x}(j+1|j) = A^*(j)\tilde{x}(j|j) + m^*(j) \simeq f[\tilde{x}(j|j), j]}$

and

$$\boxed{\begin{aligned} \Psi_{\tilde{e}(j+1|j+1)} &= \Psi_{\tilde{e}(j+1|j)} [I + C^{*T}(j+1)\Psi^{-1}_{v(j+1)} C^*(j+1)\Psi_{\tilde{e}(j+1|j)}]^{-1} \\ \Psi_{\tilde{e}(0|0)} &= \Psi_{x(0)} \end{aligned}}$$

where

$$\boxed{\Psi_{\tilde{e}(j+1|j)} = \Gamma(\Psi_{\tilde{e}(j|j)})}$$

$$= A^*(j)\Psi_{\tilde{e}(j|j)}A^{*T}(j) + B(j)\Psi_{u(j)}B^T(j)$$

An implementation on-line of $x^*(j)$

is [illegible]

$$x^*(j) = \tilde{x}(j|j-1)$$

In this case the approximated filtering algorithm is known as

EXTENDED KALMAN FILTER (EKF).

Here, we do not distinguish between deterministic and random component of $x(j)$, since in this case $x^*(j)$ is itself random.

The EKF is

$$\begin{cases} \tilde{x}(j+1|j+1) = \tilde{x}(j+1|j) + \Psi_{\tilde{e}(j+1|j+1)} C^{*T}(j+1) \cdot \\ \cdot \Psi^{-1}_{v(j+1)} \left[z(j+1) - g(\tilde{x}(j+1,j), j+1) \right] \\ \qquad\qquad j = 0, \dots, k-1 \\ \tilde{x}(0|0) = \bar{x}(0) \\ \tilde{x}(j+1|j) = f(\tilde{x}(j|j), j) \end{cases}$$

$$\begin{cases} \Psi_{\tilde{e}(j+1|j+1)} = \Psi_{\tilde{e}(j+1|j)} \left[I + C^{*T}(j+1) \Psi^{-1}_{v(j+1)} C^{*}(j+1) \Psi_{\tilde{e}(j+1|j)} \right]^{-1} \\ \qquad\qquad j = 0, \dots, k-1, \\ \Psi_{\tilde{e}(0|0)} = \Psi_{x(0)} \\ \Psi_{\tilde{e}(j+1|j)} = A^{*}(j) \Psi_{\tilde{e}(j|j)} A^{*T}(j) + B(j) \Psi_{u(j)} B^{T}(j) \end{cases}$$

and $\overset{*}{A}(j)$ and $C^{*}(j)$ are defined as usual with $x^{*}(j) = \tilde{x}(j|j-1)$.

~~STEADY STATE KALMAN FILTER~~

The KF (47) has some disadvantages. The matrices characterizing (47) are time-varying. In particular, the Kalman gain depends on $\Psi_{\tilde{e}(j+1|j+1)}$. A KF which constant matrices would be easier to implement. It is also of paramount importance that the KF be asymptotically stable, because this would leave us free to choose any initial conditions $\tilde{x}(0|0)$ for the filter, not necessarily $\tilde{x}(0|0) = E[x(0)]$

which may be unknown.

Indeed, for a stable filter the effect of initial conditions attenuates as time increases.

With these motivations, we first assume, in addition, that the system matrices are constant and $\{m(j)\}$ and $\{v(j)\}$ are i.i.d.:

$$\boxed{\begin{aligned} x(j+1) &= A\,x(j) + B\,m(j), \quad j=0,\dots,k-1, \\ z(j) &= C\,x(j) + v(j), \quad j=1,\dots,k \end{aligned}}$$

with $\Psi_{m(j)}=\Psi_m$, $\Psi_{v(j)}=\Psi_v$ $\forall j$, (64)

and look for a "steady-state" implementation of the KF (47). Clearly, the steady-state filter we obtain is no more a MMSE estimate, but it is auspicable that it will approach the MMSE estimate, given by (47), as time increases.

In order to define a "steady-state" KF, we will analyze few properties of the covariance equation in (47).

(1) We have the following cases:

$$\begin{cases} \Psi_{\tilde{e}(1|1)} \geq \Psi_{\tilde{e}(0|0)} \Rightarrow \Psi_{\tilde{e}(j+1|j+1)} \geq \Psi_{\tilde{e}(j|j)} \quad \forall j \geq 0, \\ \Psi_{\tilde{e}(1|1)} \leq \Psi_{\tilde{e}(0|0)} \Rightarrow \Psi_{\tilde{e}(j+1|j+1)} \leq \Psi_{\tilde{e}(j|j)} \quad \forall j \geq 0. \end{cases}$$

This means that the covariance $\Psi_{\tilde{e}(j|j)}$ is monotone and it is increasing (decreasing) according if it is such at the first time increment.

Next, consider the matrices Ψ_r which identically satisfy the covariance equation in (47):

$$\Psi_r = \Gamma(\Psi_r)\left[I + C^T \Psi_\nu^{-1} C \, \Gamma(\Psi_r)\right]^{-1}$$

which, recalling $\Gamma(\Psi_r') = A\Psi_r A^T + B\Psi_u B^T$, and after simple manipulations, gives

$$\boxed{\Psi_r + (\Psi_r C^T \Psi_\nu^{-1} C - I)(A\Psi_r A^T + B\Psi_u B^T) = 0 \qquad (65)}$$

This is a Riccati equation and we are interested in finding its symmetric and positive semi-definite solutions Ψ_z (as covariances).

② If (C, A) is observable (i.e. $\text{rank}\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} = n$), then, whatever is $\Psi_{\tilde{e}}(0|0)$,

$$\lim_{j \to \infty} \Psi_{\tilde{e}}(j|j) = \Psi_z$$

and Ψ_z is a symmetric and positive semidefinite solution of (65).

Therefore, if (C, A) is observable, there exists at least one (admissible) solution of (65).

3 If (C, A) is observable and (A, B) is controllable (i.e. $\operatorname{rank}[B \; AB \; \dots \; A^{n-1}B] = n$), then, whatever is $\Psi_{\tilde{e}}(0|0)$,

$$\lim_{j \to \infty} \Psi_{\tilde{e}}(j|j) = \Psi_z$$

and Ψ_z is the unique symmetric and positive semidefinite solution of (65). Moreover, the matrix

$$A_z = A - \Psi_z C^T \Psi_\nu^{-1} C A$$

has its eigenvalues inside the unit circle in $\mathbb{C}$.

If, in addition, (A, B) is controllable, then there exist only one (admissible) solution of (65) and it is "stabilizing" in the sense above.

Finally, also

> 4 Iff A has its eigenvalues inside the unit circle in $\mathbb{C}$, then we have the same conclusions as in 3.

Therefore, recalling that the ~~kalman~~ gain is

$$F(j+1) = \Psi_{\tilde{e}(j+1|j+1)} C^T \Psi_v^{-1}$$

and, if we define a "steady-state" Kalman gain F_r

$$F_r = \Psi_r C^T \Psi_v^{-1}$$

where Ψ_r is a symmetric and positive semidefinite solution (whenever it exists) of (65), if either

- (C, A) is observable and (A, B) is controllable

or

- A has its eigenvalues inside the unit circle in $\mathbb{C}$

then (as a result of 3 and 4)

$$\boxed{\lim_{j \to \infty} F(j+1) = \lim_{j \to \infty} F(j) = F_r \quad (66)}$$

and

$\boxed{A_r \text{ has its eigenvalues inside the unit circle in } \mathbb{C} \quad (67)}$

We want to prove the following result, which establishes that the solution of the steady-state KF

$$\begin{cases} x_r(j+1|j+1) = A x_r(j|j) + \\ \qquad F_r\left[z(j+1) - CA x_r(j|j) \right] \\ \qquad\qquad j = 0, .., k-1, \\ F_r = \Psi_r C^T \Psi_\nu^{-1} \end{cases} \quad (68)$$

tends (in quadratic mean) to the MMSE estimate $\tilde{x}(j|j)$ given by (47). In other words, $x_r(j|j)$ is asymptotically optimal.

> Assume that (66) and (67) hold true. For any value of $x_r(0|0)$ in (68)
>
> $$\lim_{j \to \infty} E\left[\| \tilde{x}(j|j) - x_r(j|j) \|^2 \right] = 0$$

Proof. We have

$$\begin{cases} \tilde{x}(j|j) = F(j)z(j) + (I - F(j)C)A\tilde{x}(j-1|j-1) \\ F(j) = \Psi_{\tilde{e}(j|j)} C^T \Psi_\nu^{-1} \end{cases}$$

and

$$\begin{cases} x_z(j|j) = F_z z(j) + (I - F_z C)A x_z(j-1|j-1) \\ F_z = \Psi_z C^T \Psi_\nu^{-1} \end{cases}$$

~~If~~ $\varepsilon(j) := \tilde{x}(j|j) - x_z(j|j)$, from above:

$$\varepsilon(j) = A_z \varepsilon(j-1) + (F(j) - F_z)\big(z(j) - CA\tilde{x}(j-1|j-1)\big) =$$

$$= A_z \varepsilon(j-1) + \eta(j)$$

where $\eta(j) := (F(j) - F_z)\big(z(j) - CA\tilde{x}(j-1|j-1)\big)$

$$= (F(j) - F_z)\, e(j) \quad \text{(see pg. 98)}$$

Recall that, by (48), $\eta(j)$ is uncorrelated with $z(1), .., z(j-1)$ and therefore with $\varepsilon(j-1)$, being a linear combination of $z(1), \dots, z(j-1)$ (by definition of $x_z(j|j)$ and $\tilde{x}(j|j)$). It follows

$$E[\|\varepsilon(j)\|^2] = E[\|A_z\varepsilon(j-1)\|^2] + E[\|\eta(j)\|^2] \leq \|A_z\|^2 E[\|\varepsilon(j-1)\|^2] + E[\|\eta(j)\|^2] \quad (69)$$

where $\|A_z\| = \sqrt{\max_i \lambda_i}$ ($\lambda_i, i=1,\dots,n$, are the eigenvalues of $A_z^T A_z$) and by (67)

$$\|A_z\| < 1 \quad . \quad (70)$$

Therefore, applying limit for $j \to \infty$ to (69) and using (66)

$$\lim_{j\to\infty} E[\|\varepsilon(j)\|^2] \leq \lim_{j\to\infty} \|A_z\|^2 E[\|\varepsilon(j-1)\|^2]$$

and iterating this inequality

$$\lim_{j\to\infty} E[\|\varepsilon(j)\|^2] \leq \lim_{j\to\infty} \|A_z\|^{2j} \cdot \|\varepsilon(0)\|^2$$

This with (70) implies

$$\lim_{j\to\infty} E\left[\|\varepsilon(j)\|^2\right]$$

$$= \lim_{j\to\infty} E\left[\|\hat{x}(j|j) - x_t(j|j)\|^2\right] = 0 \quad \Delta$$

PARAMETRIC IDENTIFICATION FOR LINEAR MODELS

Identification of a model for an object of which we want to study the properties is organised on two levels: first, the structure of the model is determined (structural identification) and, second, the parameters characterizing the structure of the model are estimated (parametric identification). A condition which must be satisfied is that the estimation problem for the parameters be identifible. The identifibility property is in general achieved by by increasing the number of measurements and selecting input sequences which are capable of evidentiating in a sufficiently rich way the relations between the parameters and the measurements. We distinguish identification methods as follows:

- under test input sequences (open loop): impulsive, sinusoidal, noise;
- under control input sequences (closed loop)

and for

- input-output models
- state-space models

METHODS WITH TEST INPUT SEQUENCE

We assume one input and one output. Let $\{w(i)\}$ the impulsive response sequence and assume that the system is asymptotically stable so that $w(i) \simeq 0 \ \forall i \geq N$ (settling time). Also $w(0)=0$.

1 IMPULSIVE TEST INPUT

$$m(s) = \begin{cases} \alpha > 0 & s=0 \\ 0 & s=1,2,\dots, \end{cases} \quad (71)$$

The input-output model is described by

$$y(j) = \sum_{i=1}^{N} w(i)\, m(j-i)$$

(by causality the output $y(j)$ does not depend on $m(j+1), m(j+2), \ldots,$).

The measurements are modeled as

$$z(j) = y(j) + \eta(j) \quad , j = 1, \ldots, N. \tag{72}$$

where $\{\eta(j)\}$ is a white sequence, i.i.d, with zero mean and $\sigma^2_{\eta(j)} = \sigma^2_\eta$ $\forall j \geq 1$. Setting

$$z = \begin{pmatrix} z(1) \\ \vdots \\ z(N) \end{pmatrix}, \quad \theta = \begin{pmatrix} w(1) \\ \vdots \\ w(N) \end{pmatrix}, \quad \nu = \begin{pmatrix} \eta(1) \\ \vdots \\ \eta(N) \end{pmatrix}$$

we write from (71)

$$z = \alpha \theta + \nu = C\theta + \nu$$

(since $y(j) = \alpha w(j)$ by (71))

where $C = \alpha I_{N \times N}$. Note that

$$\operatorname{rank} C = N$$

Therefore θ is identifiable with one measurement z and a Markov estimate of Θ

is (here coincides with a classical LS estimate)

$$\boxed{\tilde{\hat{\Theta}} = \frac{z}{\alpha} \quad \text{with } \Psi_{\tilde{\Theta}} = \Psi_{\tilde{e}} = \frac{\sigma_\eta^2}{\alpha^2} I_{N\times N}}$$

By increasing the number of tests we can improve the covariance of the estimation error. In this case we have after k iterations of the test

$$\begin{aligned} z(j) &= \alpha\Theta + v(j), \quad j = 1, \ldots, k, \\ &= C^{(k)}\Theta + v(j) \end{aligned}$$

being

$$\underbrace{z(j) = \begin{pmatrix} \zeta^{(j)}(1) \\ \vdots \\ \zeta^{(j)}(N) \end{pmatrix}, \quad v(j) = \begin{pmatrix} \eta^{(j)}(1) \\ \vdots \\ \eta^{(j)}(N) \end{pmatrix}}_{\text{the value of } \begin{pmatrix} \zeta(1) \\ \vdots \\ \zeta(N) \end{pmatrix} \text{ and } \begin{pmatrix} \eta(1) \\ \vdots \\ \eta(N) \end{pmatrix} \text{ at repetition } j}$$

and $$C^{(k)} = \alpha \left.\begin{pmatrix} I_{N\times N} \\ \vdots \\ I_{N\times N} \end{pmatrix}\right\} k \text{ times}$$

Since

$$\operatorname{rank} C^{(k)} = NH$$

θ is identifiable with k measurements and a Markov estimate (which coincides here with a classical LS estimate) is

$$\boxed{\begin{array}{l} \tilde{\theta}|_k = \dfrac{1}{\alpha k} \displaystyle\sum_{j=1}^{k} z(j) \\[2ex] \text{with } \Psi_{\tilde{\theta}|_k} = \Psi_{\tilde{e}|_k} = \dfrac{\sigma_\eta^2}{\alpha^2 k} I_{N\times N} \end{array}}$$

Note that $\lim_{k\to\infty} \Psi_{\tilde{e}|_k} = 0$

so that $\tilde{\theta}|_k$ is consistent. Note also that $\tilde{\theta}|_k$ is the arithmetic mean of the measurements, normalized by α.

2 SINUSOIDAL TEST INPUTS

$\alpha \sin \omega t$

Here we assume to apply an input $m(t) = \alpha \sin \omega t$ to the continuous-time system and let start the time from the settling time t_a (i.e. $t \geq -t_a$)

It is known that for $t \geq 0$

$$y(t) = \alpha \, |W(j\omega)| \sin(\omega t + \operatorname{Arg} W(j\omega))$$

where $|W(j\omega)|$ and $\operatorname{Arg} W(j\omega)$ are the module and, respectively, the phase of the Fourier transform of the impulsive response $W(t)$.

If we carry out noisy measurements of the output at discrete times $j = 1, .., k$ we have

$$z(j) = y(j) + v(j) \; ; \; j = 1, .., k,$$

and

$$y(j) = \alpha |W(j\omega)| \sin(\omega t + \operatorname{Arg} W(j\omega))$$

$$= \alpha [R(\omega) \sin \omega t + I(\omega) \cos \omega t]$$

where $R(\omega)$ and $I(\omega)$ are the real and imaginary part of $W(j\omega)$.

By taking R and I as unknown parameters

$$\Theta = \begin{pmatrix} R(\omega) \\ I(\omega) \end{pmatrix}$$

and

$$z(j) = C(j)\Theta + v(j), \quad j = 1, \dots, k$$

where $C(j) = \alpha (\sin \omega t \;\vdots\; \cos \omega t)$.

Collecting the measurements altogether

$$z^{(k)} = C^{(k)} \Theta + v^{(k)}$$

with

$$C^{(k)} = \begin{pmatrix} c(1) \\ c(2) \\ \vdots \\ c(k) \end{pmatrix}$$

To guarantee identifibility of θ we need that k is such that:

$$\operatorname{rank} C^{(k)} = p = 2 \qquad (72)$$

or in other words that there exist $i < j \leq k$, such that

$$\det \begin{pmatrix} \sin \omega i & \cos \omega i \\ \sin \omega j & \cos \omega j \end{pmatrix} \neq 0$$

i.e. $0 \neq \sin \omega i \cos \omega j - \cos \omega i \sin \omega j =$

$= \sin(\omega(i-j))$

$\Leftrightarrow \omega(i-j) \neq h 2\pi$, integer h

$\Leftrightarrow$ $\boxed{\omega \text{ is not a multiple of } 2\pi}$

Once (72) is guaranteed,
we can implement a LS (Markov) estimate of θ, with standard assumptions on $\{v(j)\}$:

$$\tilde{\theta}|_k = C^{(k)\#} z^{(k)}$$
$$\Psi_{\tilde{\theta}|_k} = \Psi_{\tilde{e}|_k} = \sigma_v^2 \left(C^{(k)T} C^{(k)} \right)^{-1} \qquad (73)$$

If Δ is the duration of the time in which the k measurements are carried out and assuming that each measurement is carried out at each time $j \frac{2\pi}{\omega}$, $j = 1, \ldots, k$, so that

$$\Delta = k \frac{2\pi}{\omega}$$

by increasing k we can approximate

(73) as follows. First of all,

$$\begin{cases} \sum_{j=1}^{k} \sin^2 \omega j \simeq \int_0^{\Delta} \sin^2 \omega t \, dt = \frac{\Delta}{2} \\ \sum_{j=1}^{k} \cos^2 \omega j \simeq \int_0^{\Delta} \cos^2 \omega t \, dt = \frac{\Delta}{2} \\ \sum_{j=1}^{k} \sin \omega j \cos \omega j \simeq \int_0^{\Delta} \sin \omega t \cos \omega t \, dt = 0 \\ \sum_{j=1}^{k} z(j) \sin \omega j \simeq \int_0^{\Delta} z(t) \sin \omega t \, dt \\ \sum_{j=1}^{k} z(j) \cos \omega j \simeq \int_0^{\Delta} z(t) \cos \omega t \, dt \end{cases}$$

therefore, (73) can be approximated as

$$\tilde{\theta}|_k \simeq \frac{2}{\alpha} \begin{pmatrix} \frac{1}{\Delta} \int_0^{\Delta} z(t) \sin \omega t \, dt \\ \frac{1}{\Delta} \int_0^{\Delta} z(t) \cos \omega t \, dt \end{pmatrix}$$

$$\Psi_{\tilde{\theta}|_k} \simeq \Psi_{\tilde{e}|_k} = \frac{2\sigma_v^2}{\alpha^2 \Delta} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and from here

$$E\left[\|\tilde{e}|_k\|^2\right] = \mathrm{Tr}\, \Psi_{\tilde{e}|k} \simeq \frac{4\sigma_\nu^2}{\alpha^2 \Delta} .$$

The LS estimate is obtained by multipling $z(t)$ by $\sin \omega t$ (resp. $\cos \omega t$) and taking the (normalized) integral over Δ.

The mean square error can be reduced by increasing the duration Δ or the amplitude of the sinusoidal input.

3 NOISY TEST INPUTS (WIENER HOPF METHOD)

Here we assume to use noise as input, in particular a sequence $\{m(j)\}$, $j = -N, -N+1, \dots$

We assume that $\{m(j)\}$ and $\{v(j)\}$ are i.i.d, white and mutually uncorrelated. One has

$$z(j) = y(j) + v(j) = \sum_{i=1}^{N} w(i)\, m(j-i)$$
$$j = 1, \dots, k > N.$$

Denote by p_{zm} the cross-correlation between $\{z(j)\}$ and $\{m(j)\}$,

$$p_{zm}(j) := E\left[z(j+\tau)\, m(\tau)\right]$$

(p_{zm} depends only on j and not separately on $j+\tau$ and τ since $\{z(j)\}$ and $\{m(j)\}$ are i.i.d.).

We have for $j = 1, \ldots, N$

$$p_{zm}(j) = E[z(j+\tau)m(\tau)\}$$

$$= \sum_{i=1}^{N} w(i) E[m(j+\tau-i)m(\tau)]$$

$$+ E[v(j+\tau)m(\tau)] = \sum_{i=1}^{N} w(i) p_m(j-i)$$

$$= w(j)\sigma_m^2$$

since $\quad p_m(j-i) = \begin{cases} \sigma_m^2 & j=i \\ 0 & i \neq j \end{cases}$.

Therefore

$$\boxed{w(j) = \frac{p_{zm}(j)}{\sigma_m^2}} \quad j = 1, \ldots, N. \tag{74}$$

The cross correlation $p_{zm}(j)$ can be approximated for $k \gg N$ as:

$$p_{zm}(j) = E[z(j+\tau)m(\tau)\} \simeq \frac{1}{k} \sum_{\tau=-j+1}^{k-j} \breve{z}(j+\tau)\breve{m}(\tau)$$

(by the weak law of large numbers A.6):

indeed, if $A_k := \frac{1}{k} \sum_{\tau=-j+1}^{k-j} z(j+\tau)m(\tau)$

then

$$A_k \xrightarrow{q.m.} E[A_k] = \frac{1}{k} \sum_{\tau=-j+1}^{k-j} E[z(j+\tau)m(\tau)]$$

$$= E[z(j+\tau)m(\tau)]$$

(since $\{z(j)\}$ and $\{m(j)\}$ are i.i.d.).

METHODS WITH CONTROL INPUT SEQUENCE

$\boxed{\text{① Identification of } \{w(i)\} \text{, with deterministic and known } \{m(j)\}.}$

Assume the system is asymptotically stable, with $w(j) \simeq 0 \ \forall j \geq N$.

We have seen that

$$z(j) = y(j) + v(j) =$$

$$= \sum_{i=1}^{N} w(i)\, m(j-i) + v(j)$$

$$= c(j)\,\theta + v(j) \qquad j=1,\dots,k,$$

where
$$c(j) = \begin{pmatrix} m(j-1) \\ \vdots \\ m(j-N) \end{pmatrix}$$

$$\theta = \begin{pmatrix} w(1) \\ \vdots \\ w(N) \end{pmatrix}$$

Assume $\{m(j)\}$ deterministic and known, $\{v(j)\}$ white, with zero mean and known $\sigma_v^2(j)$, $j=1,\dots,k$.

Collecting the measurements altogether

$$z^{(k)} = C^{(k)}\,\theta + v^{(k)}$$

~~As~~ long as

$$\boxed{\operatorname{rank} C^{(k)} = N} \tag{75}$$

we can implement a Markov estimate $\tilde{\Theta}|_k$ of Θ:

$$\begin{cases} \tilde{\Theta}|_k = \left(C^{(k)T} \Psi_{\nu(k)}^{-1} C^{(k)}\right)^{-1} C^{(k)T} \Psi_{\nu(k)}^{-1} z^{(k)} \\ \Psi_{\tilde{\Theta}|_k} = \Psi_{\tilde{e}|_k} = \left(C^{(k)T} \Psi_{\nu(k)}^{-1} C^{(k)}\right)^{-1} \end{cases} \tag{76}$$

Condition (75) is guaranteed as long as the control input sequence $\{m(j)\}$ ~~is~~ is sufficiently "rich". If we assume, in addition, that $\{\nu(j)\}$ is i.i.d. and the value $m(j)$ of $\{m(j)\}$ are the realizations $\breve{m}(j)$ of a i.i.d. random sequence $\{m(j)\}$ with zero mean and variance σ_m^2, we have the following approximations

$$C^{(k)^T} C^{(k)} =$$

$$\frac{1}{k}\begin{bmatrix} \sum_{j=1}^{k} \breve{m}^2(j-1) & \cdots\cdots & \sum_{j=1}^{k} \breve{m}(j-1)\breve{m}(j-N) \\ \vdots & & \\ \sum_{j=1}^{k} \breve{m}(j-N)\breve{m}(j-1) & \cdots\cdots & \sum_{j=1}^{k} \breve{m}^2(j-N) \end{bmatrix}$$

and (by the weak large numbers law)

$$\frac{1}{k}\sum_{j=1}^{k} \breve{m}^2(j-\tau) \simeq E[m^2(j-\tau)] = \sigma_m^2$$

$$\frac{1}{k}\sum_{j=1}^{k} \breve{m}(j-\tau)m(j-s) \simeq E[m(j-\tau)m(j-s)] = 0 \quad , \tau \neq s .$$

We obtain

$$\tilde{\Theta}\Big|_k = \begin{pmatrix} \frac{1}{\sigma_m^2}\frac{1}{k}\sum_{\tau=1}^{k} \breve{z}(\tau)\breve{m}(\tau-1) \\ \vdots \\ \frac{1}{\sigma_m^2}\frac{1}{k}\sum_{\tau=1}^{k} \breve{z}(\tau)\breve{m}(\tau-N) \end{pmatrix} \simeq \frac{1}{\sigma_m^2}\begin{pmatrix} p_{zm}(1) \\ \vdots \\ p_{zm}(N) \end{pmatrix}$$

which recovers (74).

(2) Parametric identification of $\{w(i)\}$ with deterministic and known $\{m(j)\}$

The method for estimating $\{w(i)\}$ requires asymptotic stability of the system and a very large number of parameters (as many as $w(1), \ldots, w(N)$). The impulsive response can be, alternatevely, identified through its characterising parameters, for example assuming its poles are real and distinct,

$$w(i) = \sum_{r=1}^{n} k_r \lambda_r^{i-1}, \quad i = 1, 2, \ldots$$

where λ_r are the real poles and k_r are reals. Therefore the parameter vector is

$$\Theta = \begin{pmatrix} k_1 \\ \vdots \\ k_n \\ \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

This setup allows for estimating Θ even for ~~in~~ unstable systems and reduces the number of parameters to $2n$. We have

$$z(j) = y(j) + v(j) =$$
$$= \sum_{i=1}^{\infty} w(i)\, m(j-i) + v(j)$$
$$= g(\Theta; j) + v(j) \qquad j = 1, \dots, k$$

where $g(\Theta, j) := \sum_{i=1}^{\infty} w(i)\, m(j-i)$, and

$$z^{(k)} = g^{(k)}(\theta) + v^{(k)}$$

with $g^{(k)}(\theta) = \begin{pmatrix} g(\theta,1) \\ \vdots \\ g(\theta,k) \end{pmatrix}$.

We implement a ML $\tilde{\theta}|_k$ estimate of θ, as long as $g^{(k)}(\theta)$ is invertible over $\mathbb{R}^{2n}$:

$$\tilde{\theta}|_k = \underset{\mathbb{R}^{2n}}{\operatorname{argmax}} \ln \phi_{z^{(k)}}(\breve{z}^{(k)}, \theta)$$

$\boxed{\text{3) Identification of the coefficients of the input-output model, with deterministic and known } \{m(j)\}}$

With fixed n, consider the input-output model

$$y(j) + \sum_{i=1}^{n} a_i y(j-i) = \sum_{i=1}^{n} b_i m(j-i)$$

Assume $\{v(j)\}$ white, i.i.d. with zero mean. The parameter vector is

$$\theta = \begin{pmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{pmatrix}$$

We illustrate here the procedure for estimating θ by minimizing the "measurement prediction error" (MPE). The measurement equation is

$$z(j) = c(j)\theta + v(j), \quad j = 1, \dots, k$$

with

$$c(j) := \left(-y(j-1) \,\vdots\, \dots \,-y(j-n) \,\vdots\, m(j-1) \,\vdots\, \dots \,\vdots\, m(j-n)\right).$$

Define

$$\eta(j) := \gamma(j)\,\Theta$$

where

$$\gamma(j) := \left(-\breve{z}(j-1) \,\vdots\, \cdots \,\vdots\, -\breve{z}(j-n) \,\vdots\, m(j-1) \,\vdots\, \cdots \,\vdots\, m(j-n)\right)$$

which is defined for $j = n+1, \ldots, k$.

The measurement prediction error is defined as:

$$\boxed{\varepsilon(j) := z(j) - \eta(j) = z(j) - \gamma(j)\Theta, \quad j = n+1, \ldots, k}$$

Let introduce the cost function

$$J(\Theta) = \frac{1}{2} \left\| z^{(k)} - \Gamma^{(k)}\Theta \right\|^2 \qquad (77)$$

where

$$z^{(k)} = \begin{pmatrix} z(n+1) \\ \vdots \\ z(k) \end{pmatrix}, \quad \Gamma^{(k)} = \begin{pmatrix} \gamma(n+1) \\ \vdots \\ \gamma(k) \end{pmatrix}.$$

The MPE estimate is defined as

$$\tilde{\Theta}|_k = \arg\min_{\mathbb{R}^{2n}} J(\Theta) \qquad (78)$$

We remark that the above minimization procedure does not lead to a LS estimate $\tilde{\Theta}|_k$ since $\Gamma^{(k)}$ in (77) depends on Θ through $\breve{z}(n), \ldots, \breve{z}(k-n)$. However, in order to solve (78) we proceed as for LS estimation. First of all, assume k is such that

$$\boxed{\operatorname{rank} \Gamma^{(k)} = 2n \leq k-n}$$

which is the identifiability condition for Θ.

Let's use the necessary condition for (78):

$$0 = \left.\frac{dJ}{d\theta}\right|_{\theta=\tilde{\theta}|_k} = -2\left(z^{(k)} - \Gamma^{(k)}\theta|_k\right)^T \Gamma^{(k)}$$

which leads to

$$\tilde{\theta}|_k = \Gamma^{(k)\#} z^{(k)} \qquad (79)$$

Moreover,

$$\frac{d^2 J}{d\theta^2} = \frac{d}{d\theta}\left(\frac{dJ}{d\theta}\right)^T = 2\, I_{2n\times 2n} > 0$$

which implies that $J(\theta)$ is strictly convex. Therefore, (79) is the MPE estimates.

The estimate $\tilde{\Theta}|_k$ is not a LS estimates and, as such, it is not centered. However, we can guarantee asymptotic centering and consistency as follows. Let $k \gg n$ and fix $r \geq n+1$ such that $\frac{k}{r}$ is an integer. Consider

$$\eta(jr) = \gamma(jr)\Theta \qquad j = 1, \dots, \frac{k}{r} .$$

In place of (77) consider

$$J(\Theta) = \frac{1}{2} \sum_{j=1}^{\frac{k}{r}} \varepsilon^2(jr)$$

$$= \frac{1}{2} \| \bar{z}^{(k)} - \bar{F}^{(k)}\Theta \|^2$$

where

$$\bar{z}^{(k)} = \begin{pmatrix} z(\tau) \\ \vdots \\ z(k) \end{pmatrix}, \quad \bar{\Gamma}^{(k)} = \begin{pmatrix} \gamma(\tau) \\ \vdots \\ \gamma(k) \end{pmatrix}.$$

Assuming again

$$\operatorname{rank} \bar{\Gamma}^{(k)} = 2n \leq \frac{k}{\tau}$$

(which implies $k \geq 2\tau n \geq 2(n+1)n$)

we get

$$\tilde{\theta}|_k = \underset{\mathbb{R}^{2n}}{\operatorname{argmin}}\, J(\theta) = \bar{\Gamma}^{(k)\#} \bar{z}^{(k)}$$

and

$$\tilde{e}|_k = \theta - \tilde{\theta}|_k = -\bar{\Gamma}^{(k)} \bar{\varepsilon}^{(k)}$$

where

$$\bar{\varepsilon}^{(k)} = \begin{pmatrix} \varepsilon(\tau) \\ \vdots \\ \varepsilon(k) \end{pmatrix}$$

By definition of $\varepsilon(j)$

$$\varepsilon(jr) = z(jr) - \gamma(jr)\theta =$$

$$= \big(c(jr) - \gamma(jr)\big)\theta + v(jr)$$

$$= v(jr) + a_1 v(jr-1) + \dots + a_n v(jr-n)$$

$$j = 1, \dots, \frac{k}{r}$$

From this it follows that (since $r > n$) $\varepsilon(r), \dots, \varepsilon(k)$ are i.i.d. and uncorrelated (i.e. the sequence is white and i.i.d.). If

a $x^n + a_1 x^{n-1} + \dots + a_{n-1}x + a_n = 0$ has all its roots in the unit circle in $\mathbb{C}$

b the following are well-defined

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k} m(j) = \overline{m}$$

and

$$\varphi_m(i) := \lim \frac{1}{k} \sum_{j=0}^{k} (m(j) - \bar{m})(m(j+i) - \bar{m})$$

$$i = 0, \pm 1, \pm 2, \ldots$$

$\underline{c}$ the $n \times n$ matrix

$$\Phi_m := \begin{pmatrix} \varphi_m(0) & \cdots & \varphi_m(n-1) \\ \varphi_m(-1) & \cdots & \varphi_m(n-2) \\ \vdots & \vdots & \vdots \\ \varphi_m(-n+1) & \cdots & \varphi_m(0) \end{pmatrix}$$

is nonsingular,

then

$$\lim_{k \to \infty} E\left[\| \tilde{\theta}|_k - \theta \|\right] = 0$$

which implies asymptotic centering of $\tilde{\theta}|_k$ and (by Cebishev inequality) consistency.

4 Identification of the coefficients of the state-space model, with deterministic and known $\{m(j)\}$

Recalling the input-output model

$$y(j) + \sum_{i=1}^{n} a_i y(j-i) = \sum_{i=1}^{n} b_i m(j-i) \qquad (80)$$

it is clear that if $\{A, B, C\}$ is a state-space realization of (80), then any other $\{A', B', C'\}$ such that $A' = TAT^{-1}$, $B' = TB$ and $C' = CT^{-1}$ for some nonsingular T is a state-space realization of (80). Recall that a state space (or the canonical observable form

$$A' = A^T, \; B' = C^T, \; C' = B^T$$

Both are by construction controllable and observable :

$$\text{rank} \left[B \; AB \ldots A^{n-1}B \right] = \text{rank} \left[B' \; A'B' \ldots A'^{n-1}B' \right] = n$$

and

$$\text{rank} \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} = \text{rank} \begin{bmatrix} C' \\ C'A' \\ \vdots \\ C'A'^{n-1} \end{bmatrix} = n$$

First of all, we have to guarantee that the parameter vector

$$\Theta = \begin{pmatrix} a_n \\ \vdots \\ a_1 \\ b_n \\ \vdots \\ b_1 \end{pmatrix}$$

is identifiable.

realization of (80) is a triple $\{A, B, C\}$ such that

$$w(i) = \begin{cases} C A^{i-1} B & i \geq 1 \\ 0 & i = 0 \end{cases}$$

and $w(i)$ is the impulsive respon se associated with (80) (equal to $y(i)$ when $m(i) = \begin{cases} 1 & i=0 \\ 0 & i \neq 0 \end{cases}$).

Therefore, the estimation problem for θ, when referring to the state-space model, requires a canonical state-space structure to be fixed for reference. In particular, we may refer to the canonical controllable form

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \dots & -a_1 \end{pmatrix}, B = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, C = (b_n \; b_{n-1} \dots b_2 \; b_1)$$

Recalling that

$$z(j) = y(j) + v(j), \quad j = 1, \ldots, k$$

with

$$y(j) = \sum_{i=1}^{j} w(i)\, m(j-i) := g(\theta, j) \quad (81)$$

(we are assuming here that $m(j) = 0$ $\forall\, j < 0$ and $m(0) \neq 0$ and $x(0) = 0$)

we want to show that

$$g^{(k)}(\theta) = \begin{pmatrix} g(\theta,1) \\ \vdots \\ g(\theta,k) \end{pmatrix}$$

is invertible for $k \geq 2n$. To this aim,

$$y^{(2n)} := \begin{pmatrix} y(1) \\ \vdots \\ y(2n) \end{pmatrix}, \quad w^{(2n)} := \begin{pmatrix} w(1) \\ \vdots \\ w(2n) \end{pmatrix}$$

$$M := \begin{pmatrix} m(0) & 0 & \cdots & 0 \\ m(1) & m(0) & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ m(2n-1) & m(2n-2) & \cdots & m(0) \end{pmatrix}$$

The equations (81) can be rewritten as

$$y^{(2n)} = M w^{(2n)}$$

But M is nonsingular since M is triangular and $m(0) \neq 0$. It follows

$$w^{(2n)} = M^{-1} y^{(2n)} \qquad (81)$$

Next, applying $m(j) = \begin{cases} 1 & j=0 \\ 0 & j \neq 0 \end{cases}$

we obtain $y(j) = w(j)$ so that from the input-output model

$$\left\{ \begin{array}{ll} w(1) = b_1 & (j=1) \\ w(2) + a_1 w(1) = b_2 & (j=2) \\ \vdots & \\ w(n) + a_1 w(n-1) + \ldots + a_{n-1} w(1) = b_n & \\ & (j=n) \end{array} \right. \qquad (82)$$

$$\begin{cases} w(n+1) + a_1 w(n) + \dots + a_{n-1} w(2) \\ \qquad\qquad + a_n w(1) = 0 \\ \vdots \\ w(2n) + a_1 w(2n-1) + \dots + a_{n-1} w(n+1) \\ \qquad\qquad + a_n w(n) = 0 \end{cases} \tag{83}$$

Setting

$$a = \begin{pmatrix} a_n \\ \vdots \\ a_1 \end{pmatrix}, \quad b = \begin{pmatrix} b_n \\ \vdots \\ b_1 \end{pmatrix}$$

and

$$W = \begin{pmatrix} w(1) & \dots & w(n) \\ \vdots & & \\ w(n) & \dots & w(2n-1) \end{pmatrix}, \quad w^* = \begin{pmatrix} w(n+1) \\ \vdots \\ w(2n) \end{pmatrix}$$

we obtain from (83)

$$W a = -w^* \tag{84}$$

Recalling $w(j) = \begin{cases} CA^{j-1}B & j \geq 1 \\ 0 & j = 0 \end{cases}$

then

$$W = \begin{pmatrix} CB & CAB & \dots & CA^{n-1}B \\ \vdots & & & \\ CA^{n-1}B & CA^{n-2}B & \dots & CA^{2n-2}B \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}}_{O} \underbrace{\begin{pmatrix} B & AB & \dots & A^{n-1}B \end{pmatrix}}_{R}$$

which is nonsingular, since O and R are. Therefore, from (84)

$$q = -W^{-1} w^* \qquad (85)$$

Using this in (82) we get

$$\underbrace{\begin{pmatrix} 0 & w(1) & w(2) & \dots & w(n-1) \\ 0 & 0 & w(1) & \dots & w(n-2) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & w(1) \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}}_{N} a + \underbrace{\begin{pmatrix} w(n) \\ \vdots \\ w(2) \\ w(1) \end{pmatrix}}_{c^*} = b$$

which gives

$$b = c^* - HW^{-1}w^* \qquad (86)$$

From (81), (85) and (86) it follows that the parameters $\theta = \begin{pmatrix} a \\ b \end{pmatrix}$ can be expressed as a function of $y^{(2n)}$ (i.e. $g^{(2n)}(\theta)$ is invertible over its domain).

Next, we consider the measurement equation

$$z(j) = y(j) + v(j) = \sum_{i=1}^{j} w(i)\, m(j-i) + v(j)$$

$$= \sum_{i=1}^{j} CA^{i-1}B\, m(j-i) + v(j)$$

$$= g(\theta, j) + v(j) \quad , \quad j = 1, .., k.$$

where $\theta = \begin{pmatrix} a \\ b \end{pmatrix}$ (notice that we are using the fact that $\{m(j)\}$ is deterministic and known).

A ML estimate can be implemented:

$$\tilde{\Theta}\big|_k = \underset{\Theta \in D_\Theta}{\operatorname{argmax}} \ln p_{z(k)}(\breve{z}^{(k)}, \Theta)$$

If $x(0) \neq 0$ (81) is modified as follows

$$y(j) = CA^j x(0) + \sum_{i=1}^{j} w(i) m(j-i)$$

$$= CA^j x(0) + \sum_{i=1}^{j} CA^{i-1} B m(j-i)$$

so that

$$z(j) = y(j) + v(j) =$$

$$= g(\Theta, j) + v(j)$$

where $g(\Theta, j) = CA^j x(0) + \sum_{i=1}^{j} CA^{i-1} B m(j-i)$

and $\Theta = \begin{pmatrix} x(0) \\ a \\ b \end{pmatrix}$

is the "extended" parameter vector.

5 Identification of the coefficients of the state-space nonlinear model, with deterministic and known $\{m(j)\}$

Here we consider a nonlinear model

$$\begin{cases} x(j+1) = f(x(j), m(j), a) & j = 1, .., k-1 \\ x(0) = x_0 & \\ y(j) = g(x(j), a) & \end{cases} \quad (87)$$

where $a \in \mathbb{R}^{\mu}$. From (87)

$$\begin{cases} x(j) = \varphi(x(0), a) := \varphi(\theta) \\ y(j) = g(\varphi(x(0), a), a) \end{cases}$$

$$:= \psi(x(0), a)$$

$$:= \psi(\theta)$$

with $\theta = \begin{pmatrix} x(0) \\ a \end{pmatrix}$. Again, a ML estimate can be worked out, as long as the identifiability property is guaranteed.

(6) Identification of the coefficients of the state-space (linear) model with random $\{m(j)\}$

Random $\{m(j)\}$ is a realistic assumption when inputs are uncertain/noisy data.

6.1 state-space model

Then we consider

$$x(j+1) = A(\theta)x(j) + B(\theta)m(j)$$
$$j = 0, \dots, k-1,$$

$$z(j) = C(\theta)x(j) + v(j)$$
$$j = 1, \dots, k,$$

with $\theta = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^{2n}$ (see problems 3 and 4). As usual we have

$$z(j) = C(\theta)x(j) + v(j)$$
$$= C(\theta)A^{j}(\theta)x(0) + \sum_{i=1}^{j} C(\theta)A^{i-1}(\theta)B(\theta) \cdot$$
$$\cdot\, m(j-i) + v(j), \quad j = 1, \dots, k.$$

174

The estimation problem of θ is highly nonlinear and a ML estimate is sought. However, even assuming $\{v(j)\}$ white, gaussian and i.i.d. it is not easy to evaluate $p_{z(k)}(\zeta^{(k)},\theta)$ (for example a product of the marginal $p_{z(j)}(\zeta(j),\theta) = p_{v(j)}(\zeta(j) - g(\theta,j))$, $j=1,..,k$). However, we can factorize $p_{z(k)}(\zeta^{(k)},\theta)$ as follows

$$p_{z(k)}(\zeta^{(k)},\theta) =$$

$$= \underbrace{\frac{p_{z(k)}(\zeta^{(k)},\theta)}{p_{z(k-1)}(\zeta^{(k-1)},\theta)}} \cdot \underbrace{\frac{p_{z(k-1)}(\zeta^{(k-1)},\theta)}{p_{z(k-2)}(\zeta^{(k-2)},\theta)}} \cdots \underbrace{\frac{p_{z(2)}(\zeta^{(2)},\theta)}{p_{z(1)}(\zeta_1,\theta)}} \underbrace{p_{z(1)}(\zeta_1,\theta)}$$

$$= p_{z(k)|z^{(k-1)}}(\zeta^{(k)},\theta) \cdot p_{z(k-1)|z^{(k-2)}}(\zeta^{(k-1)},\theta) \cdots p_{z(2)|z^{(1)}}(\zeta^{(2)},\theta)\, p_{z(1)}(\zeta_1,\theta)$$

Assume the standard Kalman assumptions on $x(0)$, $\{m(j)\}$ and $\{v(j)\}$ (see pg. 83). By these assumptions, all the densities $p_{z(k)|z^{(k-1)}}$, $p_{z(k-1)|z^{(k-2)}}, \dots, p_{z(2)|z(1)}$ and $p_{z(1)}$ are gaussian with mean $\alpha(j,\theta)$ and covariance $\beta(j,\theta)$, $j = 1, \dots, k$. In particular, for $j = 2, 3, \dots, k$

$$\boxed{\alpha(j,\theta)} = E[z(j) \mid z^{(j-1)}] =$$

$$= C(\theta) E[x(j) \mid z^{(j-1)}] = C(\theta)\tilde{x}(j|j-1)$$

$$\boxed{= C(\theta) A(\theta) \tilde{x}(j-1|j-1)}$$

$$\boxed{\beta(j,\theta)} = E\Big\{\big(z(j) - E[z(j)|z^{(j-1)}]\big) \cdot \big(z(j) - E[z(j)|z^{(j-1)}]\big)^T\Big\}$$

$$= E\Big\{\big(C(\theta)\tilde{e}(j|j-1) + v(j)\big)\big(C(\theta)\tilde{e}(j|j-1) + v(j)\big)^T\Big\}$$

$$= C^T(\theta)\Psi_{\tilde{e}(j|j-1)}C(\theta) + \Psi_{v(j)}$$

$$\boxed{\begin{aligned} &= C^T(\theta)A\Psi_{\tilde{e}(j-1|j-1)}A^T(\theta)C(\theta) \\ &+ C^T(\theta)B(\theta)\Psi_{u(j-1)}B^T(\theta)C(\theta) \\ &+ \Psi^2_{v(j)} \end{aligned}}$$

Moreover, for $j=1$

$$\boxed{\alpha(1,\theta)} = E[z(1)] = C(\theta)E[x(1)]$$

$$= \boxed{C(\theta)A(\theta)\overline{x}(0)}$$

$$\boxed{\beta(1,\theta)} = \Psi_{z(1)} = E\Big[\big(z(1)-E(z(1))\big)\cdot \big(z(1)-E(z(1))\big)^T\Big]$$

$$= E\Big[\big(C(\theta)x(1)+v(1)-C(\theta)A(\theta)\overline{x}(0)\big)\cdot \big(C(\theta)x(1)+v(1)-C(\theta)A(\theta)\overline{x}(0)\big)^T\Big]$$

$$= E\Big[\big(C(\theta)A(\theta)(x(0)-\overline{x}(0))+C(\theta)B(\theta)m(0)+v(1)\big)\cdot \big(C(\theta)A(\theta)(x(0)-\overline{x}(0))+C(\theta)B(\theta)m(0)+v(1)\big)^T\Big]$$

$$\boxed{\begin{aligned}&= C^T(\theta)A(\theta)\Psi_{x(0)}A^T(\theta)C(\theta)\\ &+ C^T(\theta)B(\theta)\Psi_{u(0)}B^T(\theta)C(\theta)\\ &+ \Psi_{v(1)}\end{aligned}}$$

Therefore

$$\ln p_{z(k)}(\breve{z}^{(k)},\theta) = \sum_{j=2}^{k} \ln p_{z(j)|z^{(j-1)}}(\breve{z}^{(j)},\theta) + \ln p_{z(1)}(\breve{z}(1),\theta)$$

$$= \sum_{j=2}^{k} \ln \frac{1}{\sqrt{2\pi}\,(\det\beta(j,\theta))^{1/2}} \cdot e^{-\frac{1}{2}(\breve{z}(j)-\alpha(j,\theta))^T \beta^{-1}(j,\theta)(\breve{z}(j)-\alpha(j,\theta))}$$

$$+ \ln \frac{1}{\sqrt{2\pi}\,(\det\beta(1,\theta))^{1/2}} \cdot e^{-\frac{1}{2}(\breve{z}(1)-\alpha(1,\theta))^T \beta^{-1}(1,\theta)(\breve{z}(1)-\alpha(1,\theta))}$$

$$= -\frac{1}{2}(k-1)\ln 2\pi - \frac{1}{2}\sum_{j=1}^{k} \ln \beta(j,\theta)$$

$$- \frac{1}{2}\sum_{j=1}^{k} (\breve{z}(j)-\alpha(j,\theta))^T \beta(j,\theta)(\breve{z}(j)-\alpha(j,\theta))$$

Finally, inside the expression of $\alpha(j,\theta)$, $\beta(j,\theta)$, $j=1,..,k$, we replace $\tilde{x}(j-1|j-1)$ and $\Psi_{\tilde{x}(j-1|j-1)}$ with their expressions in (47). Clearly, these expressions are functions of θ (through $A(\theta)$, $B(\theta)$ and $C(\theta)$). A ML estimate is then calculated as

$$\tilde{\theta}|_k = \underset{\theta \in D_\theta}{\arg\max}\ J(\theta)$$

$$= \underset{\theta \in D_\theta}{\arg\max}\ \ln p_{z(k)}(z^{(k)}, \theta)$$

6.2 ARMA (r, s) MODEL

(AutoRegressive with Moving Average)

$$z(j) = \sum_{i=1}^{r} a_i z(j-i) + m(j) \qquad (88)$$

$$+ \sum_{i=1}^{s} b_i m(j-i) \quad , j=1,..,k$$

with random $\{m(j)\}$

thus the parameter vector is

$$\theta = \begin{pmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^{r+s}$$

We translate (88) into

$$\begin{cases} x(j+1) = A(\theta)\,x(j) + B(\theta)\,m(j) & j = 0, 1, \ldots, k-1, \\ z(j) = C(\theta)\,x(j) + m(j) & j = 0, 1, \ldots, k, \end{cases}$$

where

$$x(j) := \big(z(j-1) \,\vdots\, \cdots \,\vdots\, z(j-r) \,\vdots\, m(j-1) \,\vdots\, \cdots \,\vdots\, m(j-s)\big)^T$$

$$A(\theta) := \begin{array}{cc} \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} & \begin{array}{l} \}\, r \\ \}\, s \end{array} \\ \begin{array}{cc} \underbrace{\quad}_{r} & \underbrace{\quad}_{s} \end{array} & \end{array} \qquad B(\theta) := \begin{array}{cc} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{array}{l} \Big]\, r \\ \Big]\, s \end{array} \end{array}$$

$$C(\theta) := \theta^T$$

and

$$A_1 = \begin{array}{c} \left.\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}\right] r \\ \underbrace{\hspace{4cm}}_{r} \end{array}$$

$$A_2 = \begin{array}{c} \left.\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}\right] s \\ \underbrace{\hspace{4cm}}_{s} \end{array}$$

Notice that the measurement noise $\{m(j)\}$ is now correlated with the state noise (actually, they have the same components). Therefore, we can proceed as in 6.1 for determining $J(\theta) = \ln p_{Z(k)}(\breve{z}^{(k)}, \theta)$ except for the expressions of $\tilde{x}(j-1|j-1)$ and $\Psi_{\tilde{e}(j-1|j-1)}$ which will be now given by (58) - (60), due to the correlation between the measurement and the state noise.