ESTIMATION THEORY

- **Estimate**

  Evaluation of unaccessible variables from directly accessible variables

  - $\underline{\text{Deterministic}}$ estimate :

    Deterministic relation between the two types of variables

  - $\underline{\text{Probabilistic}}$ estimate :

    Deterministic relation + a priori information on the noise

- **Estimation Problem**

  $\underset{\substack{\downarrow \\ \text{unaccessible}}}{x \in \mathbb{R}^n}$, $\underset{\substack{\downarrow \\ \text{accessible}}}{y \in \mathbb{R}^m}$, $A \in \mathbb{R}^{m \times n}$, $d \in \mathbb{R}^m$, $m > r$

  $$\underbrace{y = Ax}_{\text{deterministic relation}} + \underbrace{d}_{\text{noise}}$$

  $\underline{\text{Aim}}$: find evaluation $\hat{x}$ of x starting from y

  - Define an admissible set $D \subset \mathbb{R}^n$
    if $\hat{x} \in D \rightarrow$ acceptable evaluation

- **Weighted least square estimate**

  Optimal criteria : minimize $\varepsilon = y - A\hat{x}$
  given by the error $x - \hat{x}$

  define $\| v \| = \sqrt{v^T W v}$, $v \in \mathbb{R}^m$, $W > 0$ symm

  $\hat{x}$ optimal if : $\hat{x} = \underset{x \in D}{\operatorname{argmin}} \| y - Ax \|^2$

  under the following condition:

  $\partial / \partial x \| y - Ax \|^2 \big|_{x = \hat{x}} = 0 \quad \stackrel{\text{or}}{=} \quad -2A^T W (y - A\hat{x}) = 0$

If A full rank $r \rightarrow \hat{x} = A_W^+ y$, $A_W^+ = (A^T W A)^{-1} A^T W$
and as long as $\hat{x} \in D$

Note: $A_W^+ A = I$

If $W = I \rightarrow \mathcal{E}_2$ euclid. norm

- Interpretation as orthogonal projection

Consider $H = \mathbb{R}^n$, $\langle y_1, y_2 \rangle_H$, $y_1^T W y_2$, $y_1, y_2 \in H$

$M = \operatorname{Im}\{A\}$, $M \subset \mathbb{R}^n$

↳ generated by the columns of A

The unique vector $\hat{y} \in M$ is the orthogonal proj. of y on M such that:

$\|y - \hat{y}\| \leq \|y - Ax\|_H \quad \forall x \in \mathbb{R}^n$

Projection th: $\langle y - \hat{y}, Ax \rangle = 0 \quad \forall x \in \mathbb{R}^n$

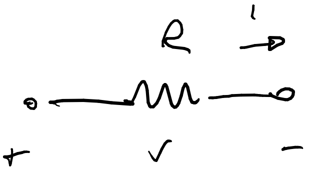
Since $\hat{y} \in M = \operatorname{Im}\{A\}$ and putting $\hat{y} = A\hat{x}$, $\hat{x} \in \mathbb{R}$

$$x^T \underbrace{A^T W (y - A\hat{x})}_{=0} = 0 \quad \forall x \in \mathbb{R}^n$$

$$\Rightarrow \hat{x} = A_W^+ y \rightarrow \hat{y} = \underbrace{A A_W^+}_{\text{orthogonal projection on } \operatorname{Im}\{A\}} y$$

~ Example:

$V = Ri$



Aim: estimate R

disturbances of the device (sensor)

If we take m measurements of (v, i) :

$$v_1 = R i_1 + d_1$$
$$\vdots$$
$$v_m = R i_m + d_m$$

If each measurement (v_i, i_i) have the same precision $\rightarrow$ least square estimate with $W = I$
(same weights for the measurements)

Estimate $\hat{R} = i^{(m)+} v^{(m)} \longrightarrow i^{(m)+} = (i^{(m)T} i^{(m)})^{-1} i^{(m)T}$

$$= \sum_{j=1}^{m} i_j v_j \Big/ \sum_{j=1}^{m} i_j^2$$

- Stochastic approach (V)

$$Y = AX + D \quad , \quad X \in \mathbb{R}^n, \; Y \in \mathbb{R}^m, \; D \in \mathbb{R}^m \quad m > n$$

random vectors — X → to be estimated, Y → measurements, D → observation noise

Available data:

$\rho_D(d)$ → density of D

$\rho_X(x)$ → density of X (if not deterministic)

First step: → evaluate density of $\rho_Y(y)$ (if X deterministic)
→ evaluate $\rho_{Y|X}(y, x)$ (if X random vector)

A) X deterministic:

Given D and Y, with $f : \mathbb{R}^m \to \mathbb{R}^m$

f invertible and differentiable over its domain

if $Y = f(D)$ then

$$\rho_Y(y) = \rho_D(f^{-1}(y)) \left| \det \frac{\partial f^{-1}(y)}{\partial y} \right|$$

Therefore $Y = AX + D = f(D)$ and

$$\rho_Y(y) = \rho_D(f^{-1}(y)) \left| \det \frac{\partial f^{-1}(y)}{\partial y} \right| = \rho_D(y - Ax) \left| \det I \right| = \underbrace{\rho_D(y - Ax)}_{\text{a priori information}}$$

since $f^{-1}(Y) = Y - AX$ and x are the values of X

Now we need to maximize $\rho_Y(y)$ by choosing a suitable $x = \hat{x}$.

$$\hat{x} \in D \subseteq \mathbb{R}^n \quad , \quad \hat{x} = \underset{x \in D \subseteq \mathbb{R}^n}{\arg\max} \; \rho_Y(y, x) \qquad y = \text{numerical values of } Y$$

↳ Maximum likelihood of X : maximizes the probability of $Y(\omega)$ to be $= y$

B) X random

$$p_{Y|X}(y,x) = \underbrace{p_D}_{\text{a priori information}}(y-Ax) \Big\} \text{a posteriori information}$$

Bayes: $$p_{X|Y}(x,y) = \frac{p_{X,Y}(x,y) \rightsquigarrow = \overbrace{p_X(x)}^{\text{a priori}}\ \overbrace{p_{Y|X}(y,x)}^{\text{a posteriori}}}{p_Y(y) \rightsquigarrow \int_{\mathbb{R}^n} p_{X,Y}(x,y)\,dx = \int_{\mathbb{R}^n} p_X(x)\, p_{Y|X}(y,x)\,dx}$$

Now choose $\hat{x}$ in order to minimize the variance of the error $x-\hat{x}$

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

and considering $\hat{x} = f(Y)$

$$\sigma^2_{x-\hat{x}} = \int_{\mathbb{R}^n} (x-f(y))^T (x-f(y))\, p_{X,Y}(x,y)\,dx\,dy$$

$f(y)$ gives the estimate of x with minimum error variance:

$$\hat{x} = f(y) = E\{X|Y\}\big|_{Y=y} = \int_{\mathbb{R}^n} x\, p_{X|Y}(x,y)\,dx$$

Estimates with minimum error variance

Given a random vector X we want to minimize the error variance

$$J(\hat{x}) = E\{\|X(\omega) - \hat{x}(\omega)\|^2\}$$

Consider only "centered" candidates of $\hat{x}(\omega)$:

$$E\{\hat{x}(\omega)\} = E\{X(\omega)\}$$

$J(\hat{x})$ is the variance of the estimation error $\mathcal{E}(\omega) = X(\omega) - \hat{x}(\omega)$

In this way, if $\hat{x}(\omega)$ is not centered:

the new estimate $\hat{x}'(\omega) = \hat{x}(\omega) + \gamma$, $\gamma = E\{X(\omega) - \hat{x}(\omega)\}$ is centered and

$$J(\hat{x}') = J(\hat{x}) - \gamma^T\gamma \leq J(\hat{x})$$

Any estimate $\hat{x}(\omega)$ is the result of a measurable function $\hat{h}$ of the measurement vector $Y(\omega)$:

$$\hat{x}(\omega) = \hat{h}(Y(\omega))$$

- Optimal problem

$$\hat{x}(\omega) = \hat{h}(Y(\omega)) = \underset{\substack{h: \mathbb{R}^m \to \mathbb{R}^n \\ \text{measurable}}}{\operatorname{argmin}} J(h(Y(\omega))) = E\{X(\omega) \mid \mathcal{Y}^Y\}$$

- Proof: Rewrite $J(\tilde{x})$, $\tilde{x} = h(Y)$, as

$$J(\tilde{x}) = E\{\|X - \tilde{x}\|^2\} = E\{\|X - \hat{x} + \hat{x} - \tilde{x}\|^2\}$$

with $\hat{x} = E\{X \mid \mathcal{Y}^Y\}$

Recall that $\langle X - \hat{x}, z\rangle_{L^2} = 0 \quad \forall\ \mathcal{Y}^Y$-measurable z (projection theorem)

But $\hat{x} - \tilde{x} = E\{X \mid \mathcal{Y}^Y\} - h(Y) = f(Y) - h(Y)$ for some $f: \mathbb{R}^m \to \mathbb{R}^n$ so that $\hat{x} - \tilde{x}$ is $\mathcal{Y}^Y$-measurable

Therefore $J(\tilde{x}) = E\{\|X - \hat{x} + \hat{x} - \tilde{x}\|^2\} =$

$$= E\{\|X - \hat{x}\|^2\} + E\{\|\hat{x} - \tilde{x}\|^2\}$$

$$\geq E\{\|X - \hat{x}\|^2\} = J(\hat{x})$$

$\hat{x} = E\{x \mid \mathcal{Y}^Y\}$ minimizes also the <u>error covariance</u> $\Psi_{\hat{\varepsilon}}$

$\hat{\varepsilon} = x - \hat{x}$

- Proof:

We have $\Psi_{\tilde{\varepsilon}} = E\{(x-\tilde{x})(x-\tilde{x})^T\}$
$= E\{(x-\hat{x}+\hat{x}-\tilde{x})(x-\hat{x}+\hat{x}-\tilde{x})^T\}$

If $\Psi' = E\{(\hat{x}-\tilde{x})(\hat{x}-\tilde{x})^T\}$, then

$$\Psi_{\tilde{\varepsilon}} = \Psi_{\hat{\varepsilon}} + \Psi' + E\{(x-\hat{x})(\hat{x}-\tilde{x})^T\} + E\{(\hat{x}-\tilde{x})(x-\hat{x})^T\}$$

But $E\{(x-\hat{x})(\hat{x}-\tilde{x})^T\} = E\{x(\hat{x}-\tilde{x})^T\} - E\{\hat{x}(\hat{x}-\tilde{x})^T\} =$

$$= E\{E\{x(\hat{x}-\tilde{x})^T \mid \mathcal{Y}^Y\}\} - E\{\hat{x}(\hat{x}-\tilde{x})^T\} =$$

$$= E\{E\{x \mid \mathcal{Y}^Y\}(\underbrace{\hat{x}-\tilde{x}}_{\mathcal{Y}^Y\text{-measurable}})^T\} - E\{\hat{x}(\hat{x}-\tilde{x})^T\} =$$

$$= E\{\hat{x}(\hat{x}-\tilde{x})^T\} - E\{\hat{x}(\hat{x}-\tilde{x})^T\} = 0$$

Therefore $\Psi_{\tilde{\varepsilon}} = \Psi_{\hat{\varepsilon}} + \Psi'$, $\Psi' \geq 0$

since where $A \geq B \to A - B \geq 0 \to \Psi_{\tilde{\varepsilon}} \geq \Psi_{\hat{\varepsilon}}$ ◀

<u>Remark</u>: $J(\tilde{x}) = \mathrm{Tr}\,\Psi_{\tilde{\varepsilon}} = \mathrm{Tr}\, E\{(x-\tilde{x})(x-\tilde{x})\} = E\{\|x-\tilde{x}\|^2\}$

(V)

Calculus of estimates with minimum variance under gaussian noise

Let $z = (X^T \; Y^T)^T$ be a gaussian vector, $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^m$

Assume $E\{z\} = m_z = \begin{pmatrix} E\{X\} \\ E\{Y\} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

covariance $\Psi_z = E\left\{ \begin{pmatrix} X \\ Y \end{pmatrix} (X^T \; Y^T) \right\} = \begin{pmatrix} \Psi_X & \Psi_{XY} \\ \Psi_{XY} & \Psi_Y \end{pmatrix}$

The estimate $\hat{X}$ with min. error variance is $\hat{X} = E\{X \mid \mathcal{Y}^Y\}$

Given the numerical values $x, \hat{x}, y$:

$$\hat{x} = E\{X \mid \mathcal{Y}^Y\}_{Y=y} = E\{X \mid Y = y\} = \int_{\mathbb{R}^n} x \, p_{X|Y}(x,y) \, dx = f(y)$$

with f measurable.

Aim: show that $\boxed{f(y) = K y}$ $\forall$ matrix K, is linear

Using Bayes theorem:

$$\hat{x} = \int_{\mathbb{R}^n} x \, p_{X|Y}(x,y) \, dx = \int_{\mathbb{R}^n} x \frac{p_{X,Y}(x,y)}{p_Y(y)} dx = \frac{1}{p_Y(y)} \int \ldots$$

$$= \frac{\int_{\mathbb{R}^n} x \, p_{X,Y}(x,y) \, dx}{\int_{\mathbb{R}^n} p_{X,Y}(x,y) \, dx}$$

$p_{X,Y}(x,y)$ gaussian $= \dfrac{1}{(2\pi)^{\frac{n+m}{2}} (\det \Psi_z)^{\frac{1}{2}}} \cdot e^{\underbrace{-\frac{1}{2} (x^T \; y^T) \Psi_z^{-1} \begin{pmatrix} x \\ y \end{pmatrix}}_{?}}$

Define $\Psi_z^{-1} = \begin{pmatrix} \bar{\Psi}_X & \bar{\Psi}_{XY} \\ \bar{\Psi}_{XY}^T & \bar{\Psi}_Y \end{pmatrix}$ with $\Psi_z \cdot \Psi_z^{-1} = 0$

$$\begin{pmatrix} \Psi_X & \Psi_{XY} \\ \Psi_{XY} & \Psi_Y \end{pmatrix} \begin{pmatrix} \bar{\Psi}_X & \bar{\Psi}_{XY} \\ \bar{\Psi}_{XY}^T & \bar{\Psi}_Y \end{pmatrix} = \begin{pmatrix} \Psi_X \bar{\Psi}_X + \Psi_{XY} \bar{\Psi}_{XY}^T & \Psi_X \bar{\Psi}_{XY} + \Psi_{XY} \bar{\Psi}_Y \\ \Psi_{YX} \bar{\Psi}_X + \Psi_Y \bar{\Psi}_{XY}^T & \Psi_{YX} \bar{\Psi}_{XY} + \Psi_Y \bar{\Psi}_Y \end{pmatrix}$$

$$= \begin{pmatrix} I_n & 0 \\ 0 & I_m \end{pmatrix} \quad (R)$$

With the previous notations:

$$(x^T y^T)\,\Psi_z^{-1}\begin{pmatrix}x\\y\end{pmatrix} = x^T\bar{\Psi}_x x + 2x^T\bar{\Psi}_{xy}\,y + y^T\bar{\Psi}_y y$$
$$= (x - My)^T\Psi_1^{-1}(x - My) + y^T\Psi_2^{-1}y \quad (*)$$

$$\Psi_1^{-1} = \bar{\Psi}_x\,,\quad M = -\bar{\Psi}_x^{-1}\bar{\Psi}_{xy}\,,\quad \Psi_2^{-1} = \bar{\Psi}_y - M^T\bar{\Psi}_x M \qquad (S)$$
$$= \bar{\Psi}_y - \bar{\Psi}_{xy}^T\bar{\Psi}_x^{-1}\bar{\Psi}_{xy}$$

From (R) $\to$ $\begin{cases}\Psi_{xy}\Psi_y^{-1} = \bar{\Psi}_x^{-1}\bar{\Psi}_{xy} & \text{2nd}\\ \bar{\Psi}_y - \bar{\Psi}_{xy}^T\bar{\Psi}_x^{-1}\bar{\Psi}_{xy} = \Psi_y^{-1} & \text{4th}\end{cases}$

(S) $\to$ $\begin{cases}\Psi_1 = \bar{\Psi}_x^{-1}\\ M = \Psi_{xy}\Psi_y^{-1}\\ \Psi_2 = \Psi_y\end{cases}$ subs in $(*)$ $\to$ $\Psi_1^{-1} = \bar{\Psi}_x$, $\Psi_2^{-1} = \Psi_y$

$P_{x,y}(x,y)$ is a gaussian density with mean My and covariance $\bar{\Psi}_x^{-1}$.

Therefore $\hat{x} = E\{X \mid Y = y\} = My = \Psi_{xy}\Psi_y^{-1}y$

From the 2nd of (R): $\bar{\Psi}_{xy} = -\bar{\Psi}_x\Psi_{xy}\Psi_y^{-1}$

replacing in the 1st: $\bar{\Psi}_x^{-1} = \Psi_x - \Psi_{xy}\Psi_y^{-1}\Psi_{xy}^T$

Estimates with minimum error variance under non-gaussian noise

X, Y non-gaussian with zero mean.

<u>Aim</u>: estimate $\tilde{X} = KY$, $K \in \mathbb{R}^{r \times m}$ for which the error variance is minimum

We can minimize the error covariance

$$J(K) = E\{(X - KY)(X - KY)^T\}$$

The problem is to find $K^* = \underset{K \in \mathbb{R}^{r \times m}}{\arg\min}\, J(K)$

We have
$$\begin{aligned} J(K) &= E\{XX^T\} - KE\{YX^T\} - E\{XY^T\}K^T + KE\{YY^T\}K^T \\ &= \Psi_X - K\Psi_{YX} - \Psi_{XY}K^T + K\Psi_Y K^T \end{aligned}$$

- Necessary conditions for K^*

Taylor of $J(K)$ around K^*:

$$\begin{aligned} J(K^* + \Delta) &= \Psi_X - (K^* + \Delta)\Psi_{YX} - \Psi_{XY}(K^* + \Delta)^T + (K^* + \Delta)\Psi_Y(K^* + \Delta)^T \\ &= J(K^*) - \Delta(-\Psi_{YX} + \Psi_Y K^*)^T + (-\Psi_{XY} + K^*\Psi_Y)\Delta^T + o(\|\Delta\|^2) \end{aligned}$$

Therefore $-\Psi_{XY} + K^*\Psi_Y = 0 \to \boxed{K^* = \Psi_{XY}\Psi_Y^{-1}}$

and
$$\begin{cases} \tilde{X} = K^* Y \\ J(K^*) = \Psi_X - K^*\Psi_{YX} = \Psi_X - \Psi_{XY}\Psi_Y^{-1}\Psi_{YX} \end{cases}$$

$\tilde{X}$ is the linear estimate which minimizes the error variance.

$$J(K) = J(K^*) + \underbrace{\tfrac{1}{2}(K - K^*)\Psi_V(K - K^*)^T}_{=0 \text{ iff } K = K^*} \quad \forall K$$