NUMERO: 2326A ANNO: 2018

A P P U N T I

STUDENTE: Caldera Filippo

MATERIA: Estimation Filtering and System Identification - Prof. Taragna

EFSI

ESTIMATION, FILTERING

AND SYSTEM IDENTIFICATION

— M. TARAGNA —

FILIPPO
CALDERA

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 3 di 152

So..

$$V_R = R\, i_R + e$$

Assuming model exist, caratterized by some parameters and noise too.

All starting by data

Model obtained is a mathematical model, ruled by parameters.
There are two main distinction categories

① DYNAMIC ← → STATIC

DYNAMIC: Present parameters depending also by the past values

$$V_R(t_0) = R i_R(t_0) + e(t_0) + R\, i_R(t_0 + t) \ldots$$

STATIC: Present parameters depends only by the present parameters

$$V_R(t_0) = R\, i_R(t_0) + e(t_0)$$

②

→ PHYSICAL MODELS

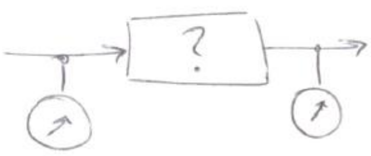
Physical laws of the system are known ⇒ possible to link mesurament to wanted parameters

⇓

SYSTEM partially known

→ BLACK-BOX MODELS

System unknown but it is tested using known input and analysing the given output.



GREY-BOX MODEL = mixture of Physical models and Black-Box

Basing of θ behavior we can classify different estimation method

$\theta(t) = \theta = \text{const}$ $\Longrightarrow$ PARAMETRIC IDENTICAL PROBLEM

- estimate indicated $\hat{\theta}$ or $\hat{\theta}_T$

- Need to find something else to something unknown.

$\theta(t) \neq \text{const}$ $\Longrightarrow$ estimator $\hat{\theta}(t|T)$, $\hat{\theta}(t|N)$

- $\hat{\theta}(t|N)$

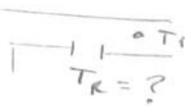
$t > t_N$ $\Longrightarrow$ PREDICTION PROBLEM

Trying to figure out what would be the value in the future

$t = t_N$ $\Longrightarrow$ FILTERING PROBLEM

Used where estimate is not measurable or is remote measure so it is estimated by measuring something else

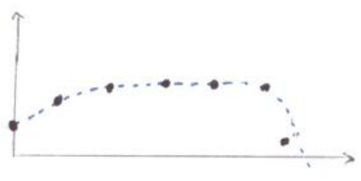
ES. Room Temperature but sensor in the hallway



$t_1 < t < t_N$ $\Longrightarrow$ REGULATION / INTERPOLATION PROBLEM

(SMOOTHING PROBLEM)

Trying to estimate the parameter value for times between the measured one



Quality of the estimation is given by how $\hat{y}(t)$ is close to $y(t)$

$\Downarrow$

Quality of estimation given by prediction error

$$\varepsilon(t) = y(t) - \hat{y}(t)$$

if $\varepsilon \to 0 \Rightarrow$ GOOD PREDICTION

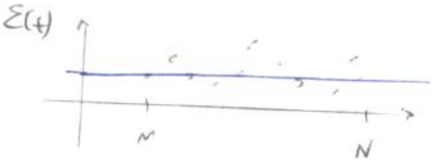
There are other additional methods like

min-MAX algorithm $\longrightarrow$ based on cost function

$$J(\vartheta) = \max |\varepsilon(x)|$$

$$\hat{\vartheta}_{optimal} = \min(J(\vartheta)) = \min\{\mathrm{MAX}\{|\varepsilon(x)|\}\}$$

Best criteria to be applyed is given by the prediction error behavior



$E[\varepsilon(t)] \neq 0$

$\Rightarrow E[\varepsilon(t)] \neq 0$ means $\varepsilon(t)$ has BIAS

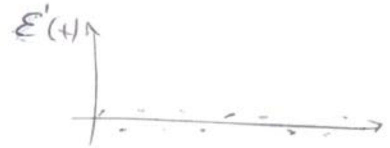
$\Downarrow$

Something in the system has been neglected

To correct the BIAS a constant term is added and considered in the estimation process

$$\hat{y}(t) = a_0 + a_1 y(t-1) + \ldots$$

$\Rightarrow$



$E[\varepsilon'(t)] = 0 \longleftarrow$ Solved BIAS

HOW TO DESCRIBE A RANDOM VARIABLE:

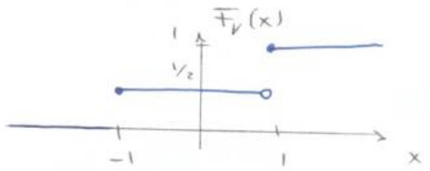
There are a large set of tools to describe a Random V.

- PROBABILITY DISTRIBUTION FUNCTION $F_V(x)$

$$F_V(x) = P(V \le x)$$

- $F(-\infty) = 0$
- $F(+\infty) = 1$
- $F(x_1) \le F(x_2) \quad \forall\, x_1 < x_2$

es.

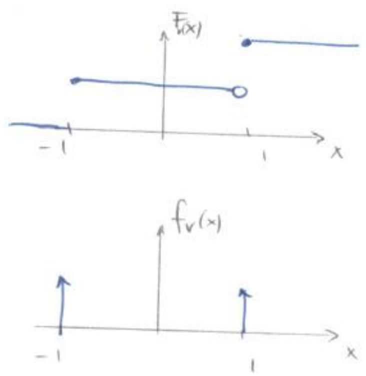


- PROBABILITY DENSITY FUNCTION $f_V(x)$

$$f_V(x) = \frac{d}{dx} F_V(x)$$

- $f(x) \ge 0 \quad \forall x$
- $\int_{-\infty}^{+\infty} f_V(x)\, dx = 1$
- $F_V(x) = \int_{-\infty}^{x} f(\xi)\, d\xi$

ES.



- K^{th} - ORDER MOMENTUM

$$m_k[V] = E[V^k] = \int_{-\infty}^{+\infty} x^k f_V(x)\, dx$$

$$m_0[V] = E[V] = 1$$

- K^{th} - ORDER CENTRAL MOMENTUM

$$\mu_k[V] = E\left[(V - E[V])^k\right] = \int_{-\infty}^{+\infty} (x - E[V])^k f_V(x)\, dx$$

$$\mu_1[V] = 0$$

$$\mu_2[V] = \sigma_V^2$$

Sometimes multitude of Random Variables are required

$\Downarrow$

Better use Random Variabiles Vector

- MEAN / EXPECTED VALUE

$$\begin{cases} E[v] = \left[E[v_1], \dots, E[v_N] \right]^T \\ E[v_i] = \int_{-\infty}^{+\infty} x_i f_i(x)\, dx \end{cases}$$

- VARIANCE MATRIX Σ_v

$$\Sigma_v = V[v] = E\left[(v - E[v])(v - E(v))^T \right] =$$

$$= \int_{\mathbb{R}^n} (x - E[v])(x - E[v])^T f(x)\, dx$$

PROPRIETIES

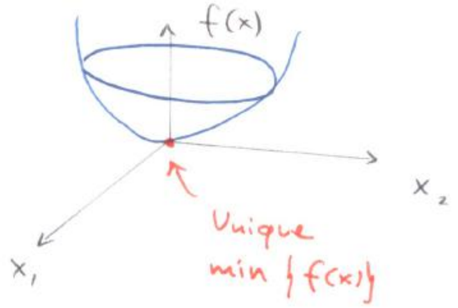
- Σ_v symmetric $\Rightarrow$ $\Sigma_v = \Sigma_v^T$
- Positive semidefinite
- $\lambda_i(\Sigma_v) = E\left[x^T (v - E[v])^2 \right] \geq 0 \quad \forall \ x \in \mathbb{R}^n$
- $[\Sigma_v]_{ii} = \sigma_i^2$ = VARIANCE
- $[\Sigma_v]_{ij} = \sigma_{ij}$ = COVARIANCE

$$\Sigma_v = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & & \vdots \\ \sigma_{21} & \sigma_2^2 & & \vdots \\ & & \ddots & \\ & & & \sigma_N^2 \end{bmatrix}$$

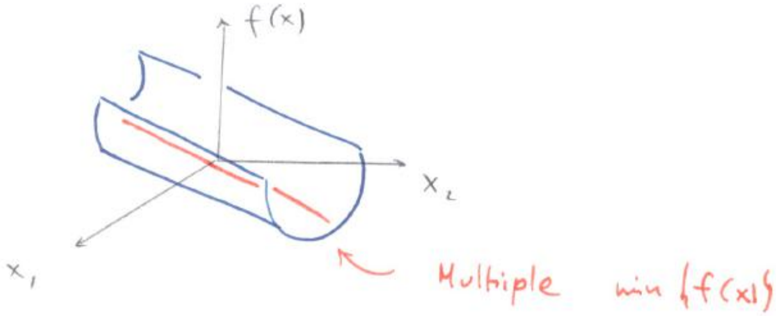
Anyway if $\det(\Sigma_v) \geq 0$ that means

- $\det(\Sigma_v) > 0 \iff \lambda_i > 0 \quad \forall i$

CASE $N=2$



- $\det(\Sigma_v) = 0 \iff \exists\, i \mid \lambda_i = 0$



if $\rho_{ij} = 0 \Rightarrow i, j$ UNCORRELATED $\not\Rightarrow$ INDIPENDENT

Having covariance matrix it's possible to "normalize it" to get NORMALIZED COVARIANCE MATRIX ρ_v

$$\rho_v = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \\ & \rho_{22} & & \\ & & \ddots & \\ & & & \rho_{NN} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \\ & 1 & & \\ & & \ddots & \\ & & & \rho_{NN} \end{bmatrix}$$

ρ_v has the same properties of Σ_v

We have a relevant case:

$$\Sigma_w = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{bmatrix} \Rightarrow \rho_w = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \\ & & & \ddots \end{bmatrix}$$

In our application $P[a<x<b]$ is

CONFIDENCE INTERVAL

=

interval where is likely to "find" our wanted values.

→ useful for estimation

Since our measurement are affected by noise (wich is considered a R.V.) then measurement itself is a R.V.

⇓

ESTIMATE IS A R.V. NOT A VALUE

⇓

To have an idea where estimate real value is located we use confidence interval.

NORMAL/GAUSSIAN USING VECTORS

if $f(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma_v}} e^{-\frac{1}{2}(x-\bar{v})^T \Sigma_v^{-1} (x-\bar{v})}$ ⇒ vector is Jointly Gaussian

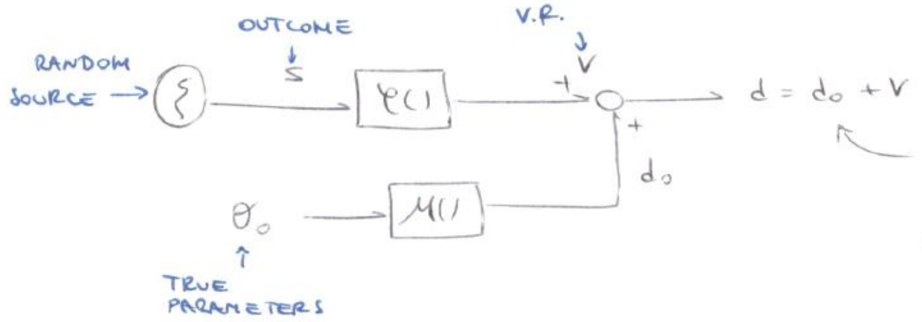
PROPRIETIES

If $v_1, \dots, v_N \sim N$ ⇒ ~~these~~ are also Jointly Normal

If
- $v_1, \dots, v_N$ Jointly Gaussian
- $v_1, \dots, v_N$ uncorrelated

⇒ $v_1, \dots, v_N$ INDIPENDENT.

PROBABILISTIC DATA DESCRIPTION



Actually this is an assumption (Noise Additive) not always have but widely used

$\hat{\theta}_0 = f(d)$ = estimation result

(since $d = d_0 + v$, v = R.V. $\Rightarrow$ d is R.V.)

In real application it's <u>not</u> sure that $\hat{\theta} \approx \theta_0$

We wish that, but how to be sure $\hat{\theta} \approx \theta_0$ if θ_0 is unknow?

Some triks needed.

=

Satisfe some estimator characteristics to be almost sure of $\theta_0 \approx \hat{\theta}$

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 23 di 152

An interesting point is ALMOST-SURE-CONVERGENCE

when N = high there are more circles to have bad datapoint.

If # bad datapoint is neglectable ⇒ Possible to neglect them

⇓

Not considering all-data point
our convergence is no more <u>sure</u>

But it's fine anyway

It's called ALMOST-SURE CONVERGENCE

After all, data point deleted are so few.

Some estimator examples

1. SAMPLE MEAN

2. :

3. WEIGHTED SAMPLE MEAN

Important fact is σ^2 can't be less than a certain value.

So, when analyzing data affected by noise is NOT possible to get noise = 0 but

$$\sigma_{\hat{\vartheta}}^2 \geq m^{-1}$$

CRAMÉR-RAO INEQUALITY

m = Fisher information quantity

depending on assumption model.

Best case possible is

$$\sigma_{\hat{\vartheta}}^2 = m^{-1}$$

In case of $\vartheta \in \mathbb{R}^n$ we have

$$Var[\hat{\vartheta}] \geq M^{-1}$$

Fisher Information Matrix

that means

$$V[\hat{\vartheta}] - M^{-1} \geq 0 \iff \forall x \in \mathbb{R}^n \quad x^T\{V[\hat{\vartheta}] - M^{-1}\}x \geq 0$$

$\Downarrow$

$$\{Var[\hat{\vartheta}] - M^{-1}\}_{ii} \geq 0 \iff V[\hat{\vartheta}] \text{ is matrix positive semi definite}$$

$$\{Var[\hat{\vartheta}]\}_{ii} - \{M^{-1}\}_{ii} \geq 0$$

$$\{Var[\hat{\vartheta}]\}_{ii} \geq \{M^{-1}\}_{ii}$$

σ^2 Fisher quantity

as scalar, σ^2 is low bounded by Fisher quantities this form $Y = \Phi \cdot \theta$ is not so true due noise

correct one is. $\underline{Y \simeq \Phi \theta}$

$Y \simeq \Phi \theta$ ← UNKNOWN ⇒ to get θ then ϕ must be invertible

⇓

- $\phi \in \mathbb{R}^{N,m}$ $(N = m)$
- $\det(\phi) \neq 0$

Not always possible

In most cases ϕ is not squared
$(N \neq m)$

we need to solve:

- find way not to get ϕ^{-1} or been able to obtain θ
- valuate how good estimation is.

} All problems solved in one shot.

Assuming to measure the distance between estimate value and the measured one.

⇓

$$\| \underbrace{\phi \theta}_{\text{output computed using } \theta} - \underbrace{Y}_{\text{measured output}} \|_2^2$$

$$\| \phi \theta - Y \|_2^2 = \sum_{k=1}^{N} \Big[\underbrace{(\phi\theta)_k}_{k^{th} \text{ row}} - \underbrace{Y_k}_{k^{th} \text{ measurement}} \Big]^2 = \sum_{k=1}^{N} \left[\varphi(k)^T \theta - y(k) \right]^2$$

need to have it as small as possible

But .. if

$$\theta = \underbrace{(\phi^T\phi)^{-1}}\phi^T Y$$

How to compute the pseudo inverse matrix of ϕ ?

3 ways

① $\theta_{LS} = pinv(\phi) * Y$

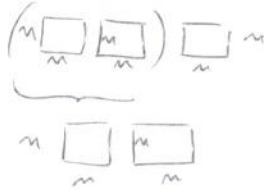
② $\theta_{LS} = inv(\underline{\phi' * \phi}) * \phi' * Y$ DO NOT USE

if $det(\phi) \to 0 \neq 0$ then θ_{LS} would be a random vector $\Rightarrow$ NO GOOD

③ $\theta_{LS} = \phi \backslash Y$

NOTE:

if ϕ squared $= N = m \Rightarrow (\phi^T\phi)^{-1}\phi^T$



if $det(\phi^T\phi) \neq 0$

$\Updownarrow$

$det(\phi) \neq 0$

$\Downarrow$

ϕ invertible itself $\Longrightarrow (\phi^T\phi)^{-1} = \phi^{-1}(\phi^T)^{-1}$

$$pinv(\phi) = (\phi^{-1})\underbrace{(\phi^T)^{-1}\phi^T}_{I} = \phi^{-1}$$

If $N = m \Rightarrow \phi$ squared $\Rightarrow pinv(\phi) = inv(\phi)$

PROBABILISTIC CHARACTERISTIC OF LS ESTIMATOR

Given $\hat{\vartheta}_{LS} = (\phi^T \phi)^{-1} \phi^T Y$

$\exists\ \hat{\vartheta}_{LS} \iff [\phi^T \phi]^{-1} \phi^T Y$ is invertible } MAIN ASSUMPTION

If $Y(t) = \varphi(t)^T \vartheta_0 + v(t)$ assuming ϑ_0 exists

$$\begin{cases} Y = \phi \vartheta_0 + V \\ V = R.V. \mid E[V] = 0 \end{cases}$$

$\Downarrow$

$$\begin{cases} \vartheta = (\phi^T \phi)^{-1} \phi^T Y \\ Y = \phi \vartheta_0 + V \end{cases} = (\phi^T \phi)^{-1} \phi^T (\phi \vartheta_0 + V) =$$

$$= (\phi^T \phi)^{-1} \phi^T \phi \vartheta_0 + (\phi^T \phi)^{-1} \phi^T V =$$

$$= \vartheta_0 + (\phi^T \phi)^{-1} \phi^T V$$

$\Downarrow$

$$E[\vartheta] = E[\vartheta_0] + \underbrace{\cancel{(\phi^T \phi)^{-1} \phi^T E[V]}}_{=0} = \vartheta_0 \Rightarrow \text{UNBIASED}$$

If $Y = \phi \vartheta_0 + V$, $V = R.V \mid E[V] = 0$ $\Rightarrow$ LS estimator UNBIASED

NOTE:

In general $\hat{\vartheta}_{LS} \neq \vartheta_0$ but

if $v \in \ker(\phi) \Rightarrow \hat{\vartheta}_{LS} = \vartheta_0$ } very rare

Then there is also another issue

$$\Sigma_v = \sigma_v^2 (\phi^T \phi)^{-1}$$

unknown in real application

$\Downarrow$

to get σ_v^2 EXPERIMENTAL NOISE VARIANCE is needed

$$\sigma^2 = \frac{J(\hat{\vartheta}_{LS})}{N-m} \underset{N \gg m}{=} \frac{J(\hat{\vartheta}_{LS})}{N} = \frac{\| Y - \phi \vartheta \|_2^2}{N}$$

To obtain a better estimation in WLS there is an additional degree of freedom given by Q.

Proven $Q = \Sigma_v^{-1} \Rightarrow V[\hat{\theta}_{WLS}] =$ min as possible

$$\underbrace{Q = \Sigma_v^{-1} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{bmatrix}^{-1} = \begin{bmatrix} 1/\sigma_1^2 & & \\ & \ddots & \\ & & \frac{1}{\sigma_N^2} \end{bmatrix}}_{\text{GAUSS-MARKOV ESTIMATE}}$$

$$\hat{\theta}_{GM} = \left[\phi^T \Sigma_v^{-1} \phi\right]^{-1} \phi^T \Sigma_v^{-1} Y$$

$$V[\hat{\theta}_{GM}] = \left[\phi^T \Sigma_v^{-1} \phi\right]^{-1}$$

Linear estimation in set membership framework
(Unknown But Bounded Framework)

Problem to be solved based in
2 kind of information

"A PRIORI" INFORMATION
= information already know before taking y.

NOISE UBB
$e \in \mathcal{B}^e$

Assumption made

$Y = \Phi \theta_o + e$

"A POSTERIORI" INFORMATION = all given by measurements

by measurements we got $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$ and

goal is to find a suitable estimator

$\hat{\theta} = \Psi(Y) \approx \theta_o$

03/23/2017

LINEAR ESTIMATION PROBLEM IN THE SET MEMBERSHIP FRAMEWORK

"A PRIORI"

$$\underbrace{Y}_{\text{Actual measurement}} = \underbrace{\phi\theta_o}_{\text{Noise Free Measurement}} + e$$

$Y \in \mathbb{R}^N$, $\theta_o \in \mathbb{R}^m$ usually $m < N$

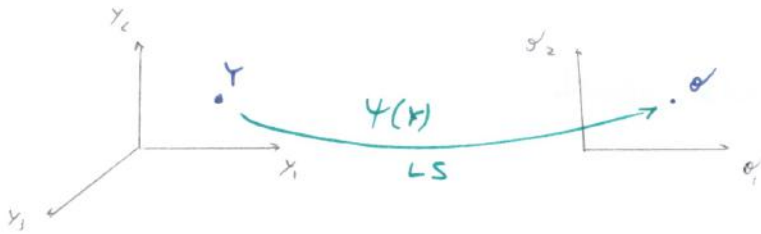
"A POSTERIORI"

All depending by Y, known the measurements

EX.

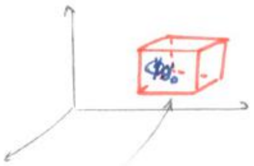
$N = 3 \implies$ Measurement Space $= \mathbb{R}^3$

$m = 2 \implies$ Parameter Space $= \mathbb{R}^2$

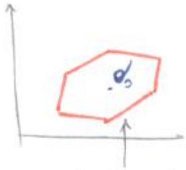


Where $\Psi(Y) \underset{LS}{=} (\phi^T\phi)^{-1}\phi^T Y = \text{pinv}(\phi)\, Y = A_{LS}\, Y$

In space is just a mapping transformation between different spaces.



cube is MUS
Measurement Uncertainty Set



MUS mapping generates
EUS
Estimate Uncertainty Set

$$MUS = y \oplus \underset{\uparrow \text{Error}}{\mathcal{B}_e^{\infty}} = \{\tilde{y} \mid y - \tilde{y} \in \mathcal{B}_e^{\infty}\}$$

⇓

FOR SURE

$$\phi\theta_0 \in MUS^{\infty}$$

Real output value is inside the cube

To prove $\theta_0 \in EUS^{\infty}$ an entire MUS mapping is required
If $\phi\theta_0$ is mapped then θ_0 is found.

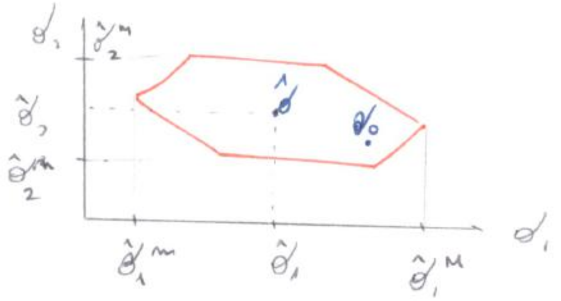
$$\left(A_{LS}\phi\theta_0 = (\phi^T\phi)^{-1}\phi^T\phi\,\theta_0 = \theta_0 \right.$$

⇓

$\theta_0 \in EUS$ FOR SURE (TOO)

Real θ_0 is inside EUS

Since, starting from Y, $\hat{\theta}$ is founded we can use ESTIMATE UNCERTAINTY INTERVALS (EUI) to better understand how θ_0 components are near to EUS border



EUI just a projection of EUS!!

Noise is UBB ⇒ Bounded by components

IDEA!
Focus on cube's vertices and mapping them.
In EUS just connect the dots

but
$\#$ vertices of MUS $= 2^N$ ⇒ If $N \uparrow$ ⇒ MAPPING WOULD BE A MESS

⇓

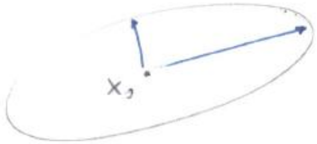
NOT USED IN PRACTICE

If Ω_x is ellipsoid centered in x_0

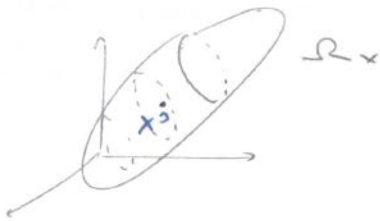
$\Downarrow$

• main axes direction given by eigenvectors of Σ_x

• semi-axes lengths given by $\sqrt{\lambda_i(\Sigma_x)}$

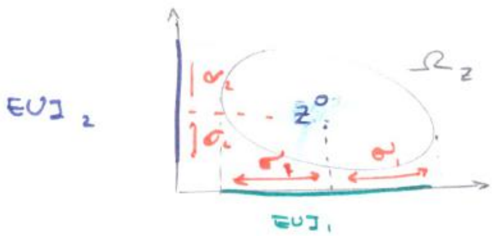


If



and $\text{rank}(A_{LS}) = m$

$\Downarrow$



$z^0 = A_{LS}\, x^0$

EUI_j evaluated using $\sigma_j = \sqrt{[(\Phi^T\Phi)^{-1}]_{jj}}$

To estimate how far y to $\phi\tilde{\theta}$ is we need to compute the distance using linear d.

$$|(y-\phi\theta)_i| = \underbrace{|y_i - \varphi_i^T\tilde{\theta}| \leq \varepsilon}$$

$$\Downarrow$$

$$-\varepsilon \leq y_i - \varphi_i^T\tilde{\theta} \leq \varepsilon$$

$$\varphi_i^T\theta \leq y_i + \varepsilon \qquad\qquad -\varphi_i^T\tilde{\theta} \leq \varepsilon - y_i$$

putting in matrices

$$\begin{bmatrix} \varphi_1^T \\ \vdots \\ \varphi_N^T \\ -\varphi_1 \\ \vdots \\ -\varphi_N \end{bmatrix} \tilde{\theta} \leq \begin{bmatrix} y_1 \\ \vdots \\ y_N \\ \vdots \\ -y_N \end{bmatrix} + \varepsilon .$$

$$\underline{H\tilde{\theta} \leq b} \longleftarrow \text{matrix linear inequalities}$$

Solution using linear programming

$$\begin{cases} \min\{c^T\tilde{\theta}\} \\ c^T = \text{Lower} \in \mathbb{R}^{1\times m} \end{cases}$$

using MATLAB

linprog$(c^T, \ldots)$

$$c^T = [0, \ldots, \underset{\substack{\uparrow \\ i^{th}}}{1}, \ldots \underset{\substack{\uparrow \\ N^H}}{0}]$$

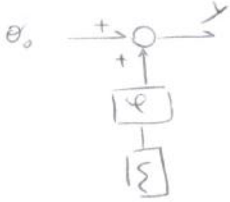
Return Parameter Uncertainty Intervals [PUI]

$$PUI = [\min \theta_j, \max(\theta_j)] = [\theta_j^m, \theta_j^M]$$

MAXIMUM LIKELIHOOD ESTIMATORS

03/30/2017

In the classical approach we had supposed θ_0 as a known function.
So, the idea was



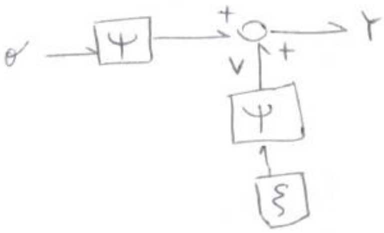
$y = \theta_0 + \phi\theta$

known

NEW APPROACH is considering θ_0 as a deterministic value instead of θ_0

$\Downarrow$

No more fixed θ_0



$y(t) = \psi(t, \theta) + v(t) = f(t, \theta)$

Random Variable

$\Downarrow$

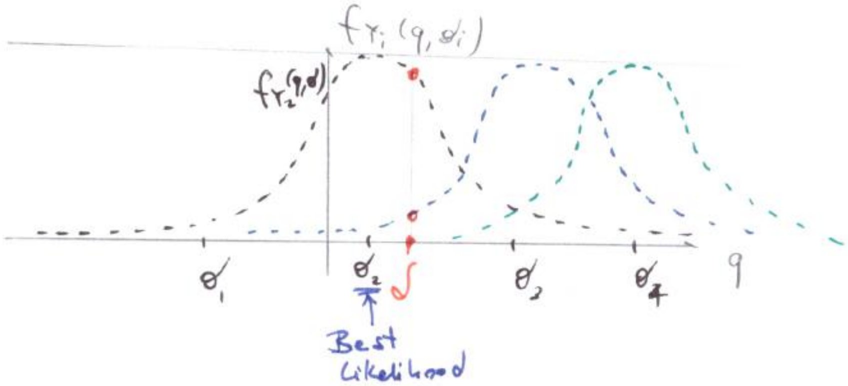
$Y(t, \theta)$ depends by

θ v

$$L(\vartheta) = f_r(q, \vartheta)\Big|_{q=\delta}$$

δ is measurement

Since δ is fixed due measure, then with multiple ϑ_i we get



Best value depends by ϑ_i, best one is ϑ_2 because is the closest one to δ and is more likely to be near the measure.

=

$f_{r_2}(q, \vartheta_2)$ is MAXIMUM LIKELIHOOD ESTIMATION

$$\vartheta_2 = \hat{\vartheta}_{ML} = \arg\{\max\{L(\vartheta_i)\}\} \quad \vartheta \in \mathbb{R}$$

→ PROPERTIES

- $E[\hat{\vartheta}_{ML}] \xrightarrow[N \to \infty]{} 0$
- $\Sigma_{\hat{\vartheta}_{ML}} \leq \Sigma_{\hat{\vartheta}} \quad N \to \infty$ } EFFICENCY Reduces uncertainties
- $\vartheta_{ML} \sim G(\ldots) \quad N \to \infty$ } Asymptotically Gaussian

⇓

If noisy measurements for $N \to \infty$ we get a Gaussian anyhow

EX. LINEAR CASE

$$Y(t,\vartheta) = \varphi'(t)\,\vartheta \Rightarrow Y(\vartheta) = \begin{bmatrix} \varphi'(1)\,\vartheta \\ \vdots \\ \varphi'(N)\,\vartheta \end{bmatrix} = \begin{bmatrix} \varphi'(1) \\ \vdots \\ \varphi'(N) \end{bmatrix} \vartheta = \Phi\vartheta$$

$$\hat{\vartheta}_{ML} = \arg_{\vartheta}\left\{ \min_{\vartheta} \left[(\underbrace{y - \Phi\vartheta}_{\text{Estimation error}})^T \Sigma_v^{-1} (\underbrace{y - \Phi\vartheta}_{\text{Estimation error}}) \right] \right\} \Rightarrow$$

(Quadratic function) Exists unique min

$$\Downarrow$$

$$\hat{\vartheta}_{ML} = \hat{\vartheta}_{GM} = \text{GAUSS-MARKOV} = \text{WEIGHTED LS}$$

NOTE!

If Φ linear $\Rightarrow \Phi$ convex $\Rightarrow \exists!\ \min(\Phi)$

If $\Sigma_v = \sigma_v^2 I_N$ indipendent, identically distributed noise $\Rightarrow \hat{\vartheta}_{ML} = \hat{\vartheta}_{LS}$

If φ LINEAR $\Longrightarrow$ GAUSS-MARKOV is

- $E[\hat{\vartheta}_{GM}] = 0$
- $\Sigma_{\hat{\vartheta}_{GM}} \leq \Sigma_{\tilde{\vartheta}} \quad \forall \tilde{\vartheta}$
- $\lim_{N\to\infty} \Sigma_{\hat{\vartheta}_{GM}} = 0$
- Gaussian

GAUSS MARKOW is BEST ONE IF ONLY IF ASSUMPTIONS ARE SATISFIED

Bayesan Estimate $= \hat{\vartheta} = E[\vartheta \mid d = \delta]$

Generic measurement — actual measurement

EX. GAUSSIAN CASE (EASIEST CASE)

$\vartheta, d \sim N(\,)$ and Jointly Gaussian

$\vartheta \in \mathbb{R}$ $\quad d \in \mathbb{R}$

$\mu_\vartheta = 0 \qquad \mu_d = 0$

then

$$\begin{bmatrix} d \\ \vartheta \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = V\begin{bmatrix} d \\ \vartheta \end{bmatrix} = \begin{bmatrix} \sigma_{dd} & \sigma_{d\vartheta} \\ \sigma_{\vartheta d} & \sigma_{\vartheta\vartheta} \end{bmatrix} \right)$$

$\sigma_{dd} = \sigma_d^2 = V[d]$

$\sigma_{\vartheta\vartheta} = \sigma_\vartheta^2 = V[\vartheta]$

$$\begin{cases} f(d, \vartheta) = c\, e^{\left\{ -\frac{1}{2} [d \ \vartheta]\, \Sigma^{-1} [d \ \vartheta]^T \right\}} \\ \Sigma = \begin{bmatrix} \sigma_{dd} & \sigma_{d\vartheta} \\ \sigma_{\vartheta d} & \sigma_{\vartheta\vartheta} \end{bmatrix} \Rightarrow \Sigma^{-1} = \frac{1}{\det \Sigma} \begin{vmatrix} \sigma_{\vartheta\vartheta} & -\sigma_{d\vartheta} \\ -\sigma_{\vartheta d} & \sigma_{dd} \end{vmatrix} \end{cases}$$

$$f(d, \vartheta) = c\, e^{\left\{ -\frac{1}{2\sigma^2} [d, \vartheta] \begin{vmatrix} \sigma_{\vartheta\vartheta}/\sigma_{dd} & -\frac{\sigma_{d\vartheta}}{\sigma_{dd}} \\ -\frac{\sigma_{\vartheta d}}{\sigma_{dd}} & 1 \end{vmatrix} \begin{vmatrix} d \\ \vartheta \end{vmatrix} \right\}}$$

$\vdots$

$$= c'\, e^{-\frac{d^2}{2\sigma_{dd}}}$$

06/06/2017

$$MSE = V\left[\vartheta - \hat{\vartheta}\right] = E\left[(\vartheta - \hat{\vartheta})^2\right] =$$

$$= E\left[\left(\vartheta - \frac{\sigma_{\vartheta d}}{\sigma_{dd}} d\right)^2\right] = E\left[\vartheta^2 - 2\frac{\sigma_{\vartheta d}}{\sigma_{dd}}\vartheta d + \frac{\sigma_{\vartheta d}^2}{\sigma_{dd}^2} d^2\right] =$$

$$= E\left[\vartheta^2\right] - 2E\left[\frac{\sigma_{\vartheta d}}{\sigma_{dd}}\vartheta d\right] + \frac{\sigma_{\vartheta d}^2}{\sigma_{dd}^2} E\left[d^2\right] =$$

$$= \sigma_{\vartheta\vartheta} - 2\frac{\sigma_{\vartheta d}^2}{\sigma_{dd}} \cdot \frac{\sigma_{\vartheta d}}{\sigma_{\vartheta d}} + \frac{\sigma_{\vartheta d}^2}{\sigma_{dd}^2}\sigma_{dd} = \sigma_{\vartheta\vartheta} - \frac{\sigma_{\vartheta d}^2}{\sigma_{dd}} = \sigma^2 > 0$$

$$=$$

$$= \sigma_{\vartheta\vartheta}\left(1 - \rho_{\vartheta d}^2\right) \geq 0$$

Mean Square Error = MSE ≥ 0

(3) d, ϑ scalar

Any p.d.f.

NON-ZERO MEAN VALUE $\mu_d = \bar{d}$, $\mu_\vartheta = \bar{\vartheta}$ $\Longrightarrow$ $\hat{\vartheta} = \bar{\vartheta} + \dfrac{\sigma_{\vartheta d}}{\sigma_{dd}}(d - \bar{d})$

$$V\left(\begin{bmatrix} d \\ \vartheta \end{bmatrix}\right) = \begin{bmatrix} \sigma_{dd} & \sigma_{\vartheta d} \\ \sigma_{d\vartheta} & \sigma_{\vartheta\vartheta} \end{bmatrix}$$

Idea is $d' = d - \bar{d}$, $\vartheta' = \vartheta - \bar{\vartheta}$

$$\sigma_{d'd'} = E\left[(d' - E[d'])^2\right] = E\left[(d')^2\right] = E\left[(d - \bar{d})^2\right] =$$
$$= E\left[d^2 - 2d\bar{d} + \bar{d}^2\right] = E[d^2] - 2E[d\bar{d}] + \bar{d}^2 =$$
$$= E[d^2] - 2\bar{d}E[d] + \bar{d}^2 = E[d^2] - 2\bar{d}\bar{d} + \bar{d}^2 =$$
$$= E[d^2] - \bar{d}^2 = E[d^2] - E[d]^2 = \sigma_{dd}$$

$\Longrightarrow \sigma_{d'd'} = \sigma_{dd}$

$$\sigma_{d'\vartheta'} = E\left[(\vartheta' - \underbrace{E[\vartheta']}_{=\bar{\vartheta}})(d' - \underbrace{E[d']}_{=\bar{d}})\right] =$$
$$= E\left[(\vartheta - \bar{\vartheta})(d - \bar{d})\right] = E\left[(\vartheta - \bar{\vartheta})(d - \bar{d})\right] = \sigma_{\vartheta d}$$

$\Longrightarrow \sigma_{d'\vartheta'} = \sigma_{d\vartheta}$

$$\begin{cases} \hat{\vartheta}' = \dfrac{\sigma_{\vartheta'd'}}{\sigma_{d'd'}} d' \\ \sigma_{d'\vartheta'} = \sigma_{d\vartheta} \\ \sigma_{d'd'} = \sigma_{dd} \end{cases} \Rightarrow \hat{\vartheta}' = \frac{\sigma_{\vartheta d}}{\sigma_{dd}}(d - \bar{d}) = (\hat{\vartheta} - \bar{\vartheta})$$

$\Downarrow$

$$\hat{\vartheta} = \bar{\vartheta} + \frac{\sigma_{\vartheta d}}{\sigma_{dd}}(d - \bar{d})$$

2

$$\hat{\vartheta} = \hat{\vartheta}_{POST} = \hat{\vartheta}_{PRIOR} + \frac{\sigma_{\vartheta d}}{\sigma_{dd}} (d - \bar{d})$$

If d affected by high uncertainties $= \sigma_{dd}$ high $\Rightarrow \hat{\vartheta}$ depends mostly by $\hat{\vartheta}_{PRIOR}$

If σ_{dd} small $\Longrightarrow \hat{\vartheta}$ is corrected by $\frac{\sigma_{\vartheta d}}{\sigma_{dd}}(d-\bar{d})$ terms

3

Posteriori uncertainty = Estimator error variance

$$V[\vartheta - \hat{\vartheta}] = E[(\vartheta - \hat{\vartheta})^2] = \sigma_{\vartheta\vartheta} - \frac{\sigma_{\vartheta d}^2}{\sigma_{dd}} = \sigma_{\vartheta\vartheta}\left(1 - \frac{\sigma_{\vartheta d}^2}{\sigma_{\vartheta\vartheta}\sigma_{dd}}\right) =$$

$$= \sigma_{\vartheta\vartheta}(1 - \rho^2)$$

if $\rho = 0 \Rightarrow \vartheta, d$ uncorrelated $\Rightarrow \hat{\vartheta} = \hat{\vartheta}_{PRIOR}$

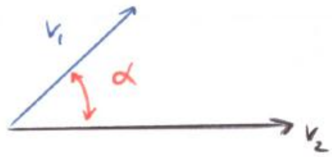
if $\rho \neq 0 \Rightarrow$ posteriori uncertainty < priori uncertainty

$$\langle v_1, v_2 \rangle = \|v_1\| \cdot \|v_2\| \cos\alpha$$

$$\Downarrow$$

$$\cos\alpha = \frac{\langle v_1, v_2 \rangle}{\|v_1\| \, \|v_2\|} = \frac{E[v_1 v_2]}{\sqrt{V[v_1] \cdot V[v_2]}} = \rho$$

$$\rho = \cos\alpha$$



$$\Downarrow$$

$$\rho = 0 \Rightarrow \alpha = \pm 90^\circ \Rightarrow v_1 \perp v_2$$

$=$

v_1, v_2 uncorrelated

$$\rho = \pm 1 \Rightarrow \alpha = 0 \Rightarrow v_1 \parallel v_2$$

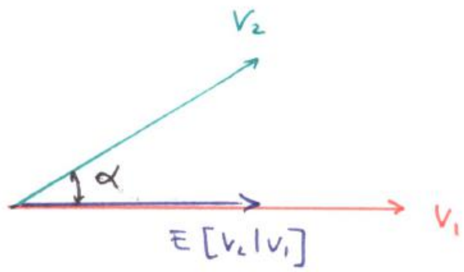
Considering scalar Gaussian case Bayes estimate is

$$\hat{v}_2 = E[v_2 | v_1] = \frac{\sigma_{21}}{\sigma_{11}} v_1 = \frac{E[v_1 v_2]}{V[v_1]} v_1$$

$$\hat{v}_2 = \frac{E[v_1 v_2]}{V[v_1]} v_1 = \frac{\langle v_1, v_2 \rangle}{\|v_1\|^2} v_1 = \frac{1}{\|v_1\|} \frac{\langle v_1, v_2 \rangle}{\|v_1\| \, \|v_2\|} \|v_2\| \, v_1 =$$

$$= \|v_2\| \underbrace{\frac{v_1}{\|v_1\|}}_{\text{versor}} \cdot \cos\alpha$$

$\Longrightarrow$ $\hat{v}_2$ has same direction of v_1



04/10/2017

Given ϑ, $d(1)$, $d(2)$ scalar R.V. where

$$\mu_{\vartheta} = \mu_{d_1} = \mu_{d_2} = 0$$

$$\Downarrow$$

$$\begin{bmatrix} \vartheta \\ d(1) \\ d(2) \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, V\begin{bmatrix} \vartheta \\ d(1) \\ d(2) \end{bmatrix} \right)$$ where

$$V\begin{bmatrix} \vartheta \\ d(1) \\ d(2) \end{bmatrix} = \Sigma = \left[\begin{array}{c|cc} \sigma_{\vartheta\vartheta} & \sigma_{\vartheta 1} & \sigma_{\vartheta 2} \\ \hline \sigma_{1\vartheta} & \sigma_{11} & \sigma_{12} \\ \sigma_{2\vartheta} & \sigma_{21} & \sigma_{22} \end{array}\right] = \left[\begin{array}{c|c} \Sigma_{\vartheta\vartheta} & \Sigma_{\vartheta d} \\ \hline \Sigma_{d\vartheta} & \Sigma_{dd} \end{array}\right]$$

Posteriori Info

$\Sigma_{\vartheta d}$ symetical $\Rightarrow \Sigma_{\vartheta\vartheta} = \Sigma_{\vartheta\vartheta}^T$

Σ_{dd} symetical $\Rightarrow \Sigma_{dd} = \Sigma_{dd}^T$

$\Sigma_{d\vartheta} = \Sigma_{\vartheta d}^T$

$$\Downarrow$$

$$\begin{bmatrix} \vartheta \\ d(1) \\ d(2) \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\vartheta\vartheta} & \Sigma_{\vartheta d} \\ \Sigma_{d\vartheta} & \Sigma_{dd} \end{bmatrix} \right)$$

So...

$$\begin{cases} E[\theta \mid d(1), d(2)] = \Sigma_{\theta d}\, \Sigma_{dd}^{-1}\, d \\ \Sigma_{\theta d} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix} \\ \Sigma_{dd}^{-1} = \dfrac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \\ d = \begin{bmatrix} d(1) \\ d(2) \end{bmatrix} \end{cases}$$

$$E[\theta \mid d(1), d(2)] = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix} \frac{1}{\sigma^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} d(1) \\ d(2) \end{bmatrix} =$$

$$= \frac{1}{\sigma^2} \begin{bmatrix} \sigma_{\theta 1} \dfrac{\sigma_{22}}{\sigma_{11}} - \sigma_{\theta 2} \dfrac{\sigma_{21}}{\sigma_{11}} & \sigma_{\theta 2} - \sigma_{\theta 1} \dfrac{\sigma_{12}}{\sigma_{11}} \end{bmatrix} \begin{bmatrix} d(1) \\ d(2) \end{bmatrix} =$$

$$= \frac{1}{\sigma^2} \left(\sigma_{\theta 1} \frac{\sigma_{22}}{\sigma_{11}} - \sigma_{\theta 2} \frac{\sigma_{21}}{\sigma_{11}} \right) d(1) + \frac{1}{\sigma^2} \left(\sigma_{\theta 2} - \sigma_{\theta 1} \frac{\sigma_{12}}{\sigma_{11}} \right) d(2)$$

where

$$\sigma^2 = \sigma_{22} - \frac{\sigma_{21}^2}{\sigma_{11}}$$

Then, using innovation ...

$$E[\vartheta \mid d(1), d(2)] = E[\vartheta \mid d(1)] + \underbrace{\frac{1}{\sigma^2}}_{=\frac{1}{\sigma_{ee}}} \underbrace{\left(\sigma_{\vartheta 2} - \sigma_{\vartheta 1} \frac{\sigma_{21}}{\sigma_{11}}\right)}_{=\sigma_{\vartheta e}} \underbrace{\Big[d(2) - E[d(2) \mid d(1)] \Big]}_{=e} =$$

$$= E[\vartheta \mid d(1)] + \frac{\sigma_{\vartheta e}}{\sigma_{ee}} e =$$

$$= E[\vartheta \mid d(1)] + E[\vartheta \mid e] = E[\vartheta \mid d(1)] + E[\vartheta \mid d(2)]$$

$$\Downarrow$$

$$\boxed{E[\vartheta \mid d(1), d(2)] = E[\vartheta \mid d(1)] + E[\vartheta \mid d(2)]}$$

Given a control law eq

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases}$$

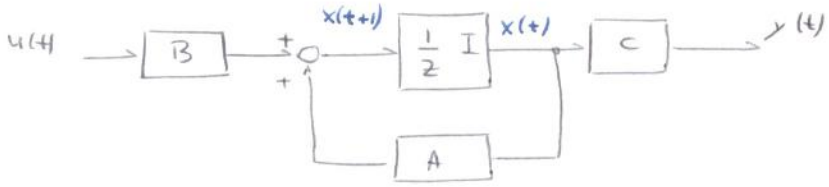
$D = 0$ $\forall$ system physically realizable

we have:

- A, B, C known
- y measured (assumed without noise)
- $x(t+1)$ unknown

Goal is to estimate $\hat{x}(t) \approx x(t)$ $\forall t$

Then, using $\mathcal{Z}$ transform and diagram blocks

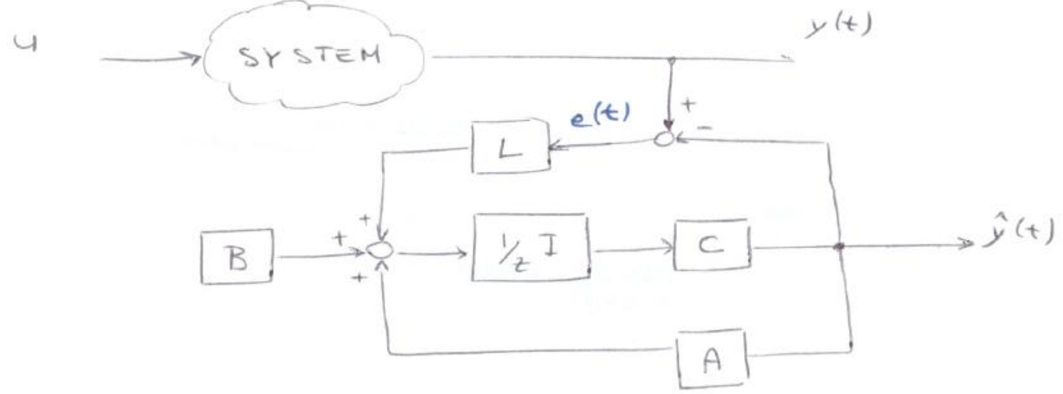


this is the system where $\hat{x}(t)$ is to be estimated

How to estimate $\hat{x}(t)$?

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 71 di 152

IDEA 2 :



NOTE
$e(t) = y(t) - \hat{y}(t)$ has the same function as bayesian estimate

So.. According to block diagram

$$\tilde{x}(t+1) = (A - LC)\,\tilde{x}(t)$$

So, now, to guarantee $\tilde{x}(t) \to 0$ is needed only

$$\left| \lambda_i \{(A - LC)\} \right| < 1 \quad \forall i$$

$\Longrightarrow$ Possible to act on L to fix the eigenvalues

$\Downarrow$

All system possible can be estimated

Architecture: use L block to bypass $|\lambda_i\{A\}| < 1$

If system fully observable $\Longrightarrow$ $\rho(M_o) = n$, $M_o = [C \;\; CA \;\; CA^2 \dots]^T$

L matrix can be designed to get

$$|\lambda_i(A - LC)| < 1 \quad \forall i$$

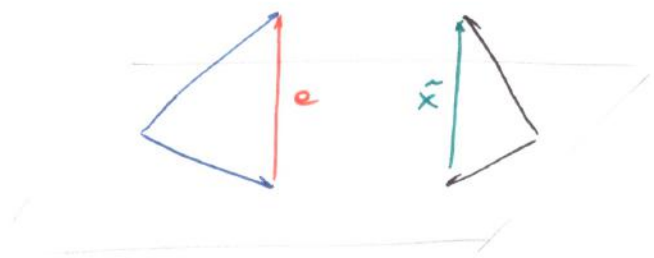
NOTE 2 $\bar{v}_1 = 0, \ \bar{v}_2 = 0 \quad \Rightarrow \quad \bar{x}(t) = 0, \ \bar{y}(t) = 0 \quad \forall t$

$$\underbrace{\hat{x}(N+r|N) = E\left[x(N+r) \mid y^N\right] = \Sigma_{x(N+r)y^N}}_{\text{Bayes Application}} \Sigma^{-1}_{y^N y^N} y^N$$

inverting matrix ⇓ Heavy task

/

Considering e and $\tilde{x}$ on the space:



$e \perp \mathcal{H}$, $\tilde{x} \perp \mathcal{H}$ $\Rightarrow$ $e \parallel \tilde{x}$ $\Rightarrow$ $e, \tilde{x}$ linearly dependent.

Optimal estimate for $x(N+1)$ based on $y^N = \begin{bmatrix} y^N \\ y^{N-1} \\ \vdots \\ y^1 \end{bmatrix}$ is:

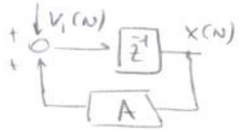
$$\hat{x}(N+1|N) = E[x(N+1)|y^N] =$$

$$= E[x(N+1)|y^{N-1}, y^N] =$$

$$\underset{\substack{\uparrow \\ \text{BAYES ESTIMATE}}}{=} E[x(N+1)|y^{N-1}] + E[x(N+1)|e(N)]$$

Then

- $E[x(N+1)|y^{N-1}] = E[Ax(N) + v_1(N)|y^{N-1}] =$



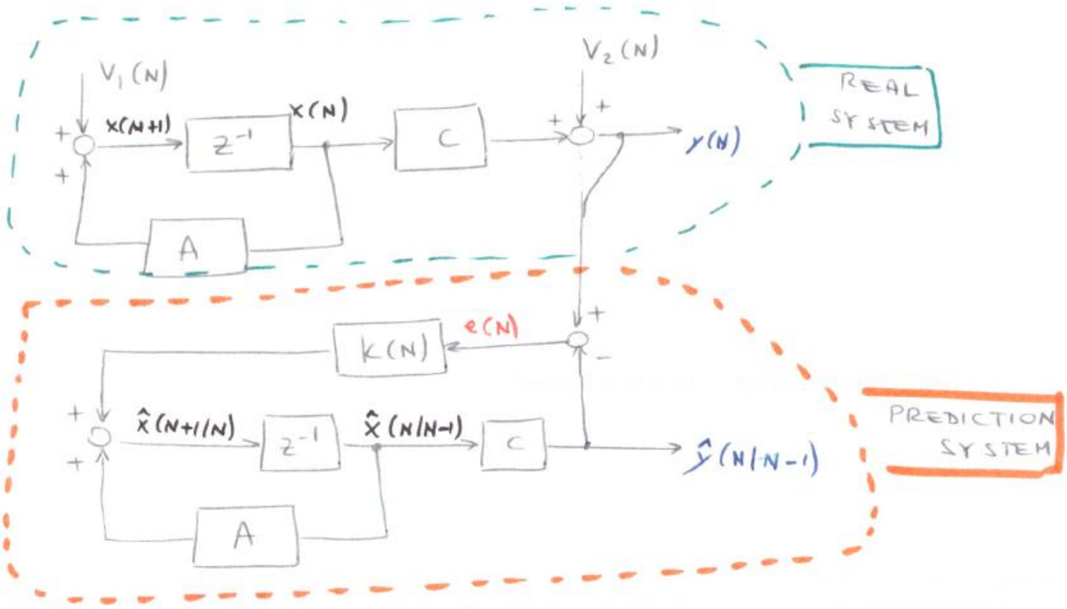
using Block Diagram

$$= A\,E[x(N)|y^{N-1}] + \underbrace{E[v_1(N)|y^{N-1}]}_{=0} =$$

$$= A\,E[x(N)|y^{N-1}] = A\,\hat{x}[N|N-1]$$

- $E[x(N+1)|e(N)] = \underbrace{\Sigma_{x(N+1)e(N)}\,\Sigma^{-1}_{e(N)e(N)}}_{K(N)}\,e(N) = K(N)\,e(N)$

Then a block diagram could be:



REMARKS

- $K(N)$ involves prediction error variable $P(N)$ which changes during time

  ⇓

  $K(N)$ is time variant.

- KALMAN is the optimal linear prediction (basing itself on BAYES)

- If
  - V_1, V_2 individually and jointly Gaussian $\forall t$
  - $x(t=1) \sim$ Gaussian
  - S linear

  $\Rightarrow$ $x(N+1)$, y^N jointly G. and individually G.

  ⇒ Bayesian Estimate of $x(N+1)$ is optimal

  ⇓

  KALMAN predictor is the best

MULTI STEP KALMAN PREDICTOR

↓

used for $x(N+r)$, $r>1$

Given the system as before the best optimal estimator for $x(N+r)$ given y^N is:

$$\hat{x}(N+r|N) = E[x(N+r)|y^N] =$$

$$= E[Ax(N+r-1) + v_1(N+r-1)|y^N] =$$

$$= A\,E[x(N+r-1)|y^N] + \underbrace{E[v_1(N+r-1)|y^N]}_{=0} =$$

$$= A\,\hat{x}(N+r-1|N)$$

due white noise

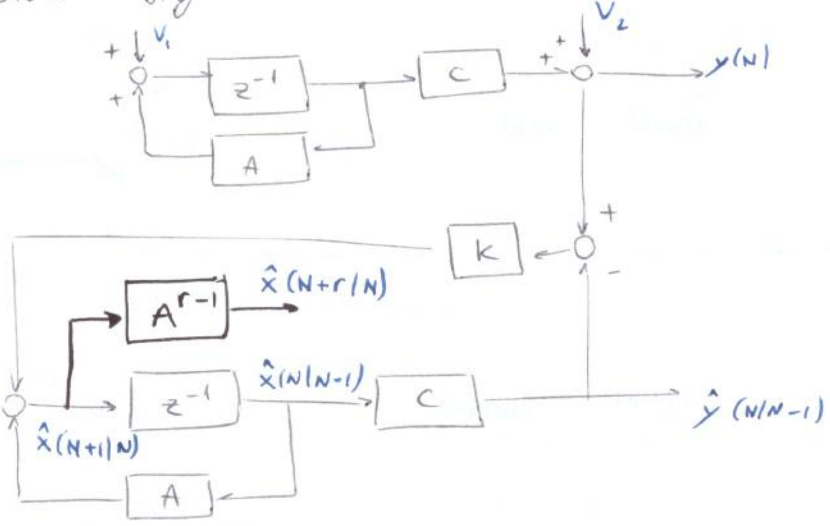
if

$$\begin{cases} \hat{x}(N+r+1|N) = A\,\hat{x}(N+r|N) \\ \hat{x}(N+r|N) = A\,\hat{x}(N+r-1|N) \quad \dots \end{cases} \Rightarrow \hat{x}(N+r|N) = A^{r-1}\,\hat{x}(N+1|N)$$

$$\hat{x}(N+r|N) = A^{r-1}\,\hat{x}(N+1|N)$$

Solution is based on the next step KALMAN Prediction

In block diagram



05/04/2017

KALMAN FILTER

=

Filtering data given measurements

IDEA

Compute filtered data as a prediction summed something else

Filtering = prediction + ?

Discovering ?

Let's assume a LTI system S

$$S: \begin{cases} x(t+1) = Ax(t) + v_1(t) \\ y(t) = Cx(t) + v_2(t) \end{cases}$$

where $v_1 \sim WN(0, V_1)$, $v_2 \sim WN(0, V_2)$, $x(t=1) \sim (\bar{x}_1, P_1)$

Goal is estimate $x(N)$ as $\hat{x}(N|N)$

Optimal estimate is $\hat{x}(N|N) = E[x(N)/y^N] = E[x(N)|y^{N-1}, y(N)]$

Applying Bayes Rewriting estimate

$$\begin{cases} \hat{x}(N|N) = E[x(N)|y^{N-1}] + E[x(N)|e(N)] - E[x(N)] \\ E[x(N)|y^{N-1}] = \hat{x}(N|N-1) \\ E[x(N)|e(N)] = \Sigma_{x(N)e(N)} \Sigma^{-1}_{e(N)e(N)} e(N) \end{cases}$$

innovation
$e(N) = y(N) - \hat{y}(N|N-1)$

It's ok but should be better to have KALMAN parameters in order to have

Filtering = Prediction + ?

$$V[\tilde{\hat{x}}(N)] = V[x(N)] - V[\tilde{x}(N|N-1)] =$$

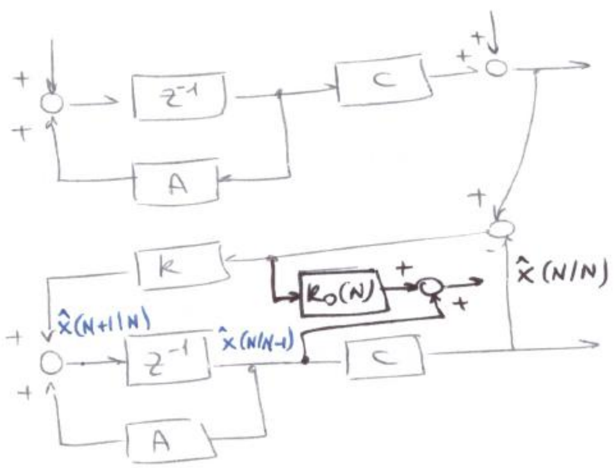
$$= P(N) - P(N)C^T[CP(N)C^T + V_2]^{-1}CP(N)$$

$$\Downarrow$$

$$V[x(N) - \tilde{x}(N|N)] \leq P[N] = V[x(N)]$$

Variance is getting smaller and smaller
=
Filtering would be more accurate.

Filtering in block diagram



PREDICTOR / CORRECTOR ONE - STEP K.P.

=

provides reliable formulation to perform the $\hat{x}(N|N-1)$ to $\hat{x}(N+1|N)$ transitions.

⇓

During the prediction is also taking place a correction

performed in 2 steps

STEP ① : KALMAN FILTER

$$\hat{x}(N|N) = \hat{x}(N|N-1) + K_0(N)\,e(N)$$

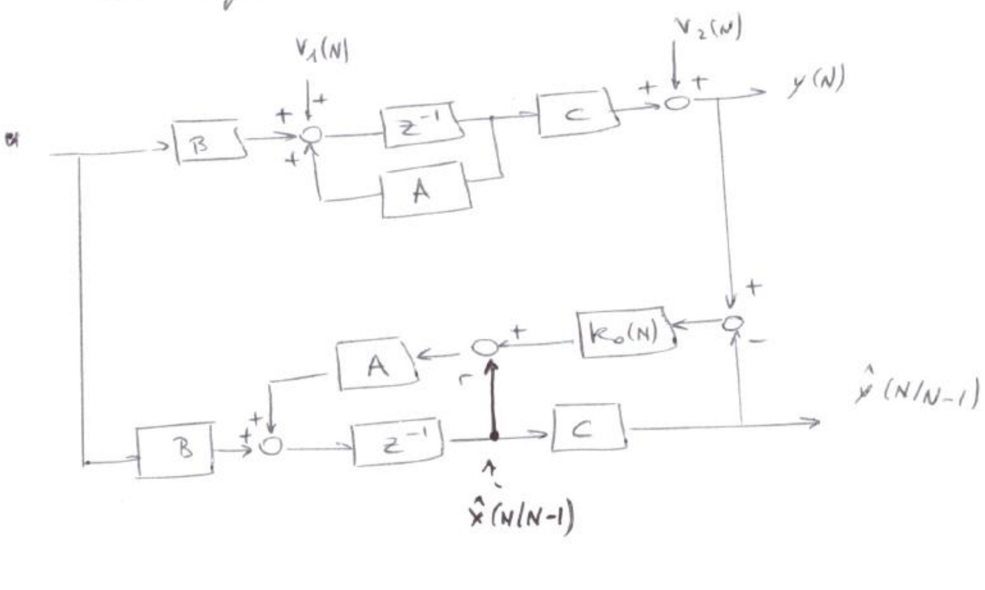
STEP ② : COMPUTING $\hat{x}(N+1|N)$ using $\hat{x}(N|N)$

$$\hat{x}(N+1|N) = A\hat{x}(N|N) + Bu(N)$$

execution in time

$$K_0(N) = P(N)C^T\left[CP(N)C^T + V_2\right]^{-1}$$
$$P_0(N) = \left[I_m - K_0(N)C\right]P(N)$$
$$e(N) = y(N) - C\hat{x}(N|N-1)$$
$$\hat{x}(N|N) = \hat{x}(N|N-1) + K_0(N)e(N)$$
$$P(N+1) = AP_0(N)A^T + V_1$$
$$X(N+1|N) = A\hat{x}(N|N) + Bu(N)$$

In block diagram



© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55 - Torino / Pagina 87 di 152

prediction error : $\tilde{x}(N+1) = x(N+1) - \hat{x}(N+1|N)$

$x(N+1) = A\, x(N) + v_1(N)$

$$\hat{x}(N+1|N) = A\,\hat{x}(N|N-1) + \bar{K}\, e(N) =$$
$$= A\,\hat{x}(N|N-1) + \bar{K}\left[y(N) - \hat{y}(N|N-1)\right] =$$
$$= A\,\hat{x}(N|N-1) + \bar{K}\left[C x(N) + v_2(N) - C\hat{x}(N|N-1)\right] =$$
$$= (A - \bar{K}C)\,\hat{x}(N|N-1) + \bar{K}C x(N) + \bar{K} v_2(N)$$

$$\tilde{x}(N+1) = A x(N) + v_1(N) - (A - \bar{K}C)\,\hat{x}(N|N-1) - \bar{K}C x(N) - \bar{K} v_2(N) =$$
$$= (A - \bar{K}C)\, x(N) + v_1(N) - (A - \bar{K}C)\,\hat{x}(N|N-1) - \bar{K} v_2(N) =$$
$$= (A - \bar{K}C)\,[\underbrace{x(N) - \hat{x}(N|N-1)}_{=\tilde{x}(N)}] + v_1(N) - \bar{K} v_2(N) =$$
$$\tilde{x}(N+1) = (A - \bar{K}C)\,\tilde{x}(N) + v_1(N) - \bar{K} v_2(N)$$

$\Downarrow$

$$\tilde{x}(N+1) = (A - \bar{K}C)\,\tilde{x}(N) + v_1(N) - \bar{K} v_2(N)$$

diff. eq. returning behavior of $\tilde{x}$

prediction error @ $N+1$ — prediction error @ N

Considering a free-noise environment $\Rightarrow v_1 = v_2 = 0 \quad \forall i$

$$\tilde{x}(N+1) = (A - \bar{K}C)\,\tilde{x}(N)$$

computing a recursive formula $\tilde{x}(N) = (A - \bar{K}C)^{N-1}\,\tilde{x}(1)$

we want an error $\tilde{x}(N)$ getting smaller and smaller

$\Updownarrow$

$$|\lambda_i(A - \bar{K}C)| < 1 \quad \forall \lambda_i$$

looks like the internal stability

NONLINEAR KALMAN FILTERING

05/11/2017

Given S a non linear system

$$S: \begin{cases} x(t+1) = f(t, x(t), u(t)) \\ y(t) = h(t, x(t), u(t)) \end{cases}$$

where f, h non linear functions, known.

It's set a nominal movement $\bar{x}(.)$, considered as a standard behavior obtained setting a nominal input $\bar{u}(.)$.
If $\bar{x}(.)$ is considered as a reference for other movements then can be defined:

comparison between theoretical movements and real ones

- State perturbation $\delta x(t) = x(t) - \bar{x}(t)$
- Input perturbation $\delta u(t) = u(t) - \bar{u}(t)$
- Output perturbation $\delta y(t) = y(t) - \bar{y}(t)$

Exploiting perturbations is possible to linearize the system S as:

$$S_L: \begin{cases} \delta x(t+1) = A(t)\,\delta x(t) + B(t)\,\delta u(t) \\ \delta y(t) = C(t)\,\delta x(t) + D(t)\,\delta u(t) \end{cases}$$

Of course δ must be relatively small to provide a good linear approximation

$$A(t) = \left.\frac{\partial f}{\partial x}\right|_{\substack{x(t)=\bar{x}(t)\\ u(t)=\bar{u}(t)}}$$

$$B(t) = \left.\frac{\partial f}{\partial u}\right|_{\substack{x(t)=\bar{x}(t)\\ u(t)=\bar{u}(t)}}$$

$$C(t) = \left.\frac{\partial h}{\partial x}\right|_{\substack{u(t)=\bar{x}(t)\\ u(t)=\bar{u}(t)}}$$

$$D(t) = \left.\frac{\partial h}{\partial u}\right|_{\substack{x(t)=\bar{x}(t)\\ u(t)=\bar{u}(t)}}$$

Jacobian Matrices

If A, B, C, D constant $\Rightarrow$ System is in equilibrium

Since the S_L is linear is possible to apply Kalman predictor

$$\bar{K}: \begin{cases} \delta\hat{x}(N+1|N) = \bar{A}(N)\,\delta\hat{x}(N|N-1) + \bar{K}(N)e(N) \\ \delta\hat{y}(N|N-1) = \bar{C}(N)\,\delta\hat{x}(N|N-1) \\ e(N) = \delta y(N) - \delta\hat{y}(N|N-1) \end{cases}$$

where

$$\begin{cases} \bar{K}(N) = \bar{A}(N)P(N)\bar{C}(N)^T\left[\bar{C}(N)P(N)\bar{C}(N)^T + V_2\right]^{-1} \\ P(N+1) = \bar{A}(N)P(N)\bar{A}(N)^T + V_1 - \bar{K}(N)\left[\bar{C}(N)P(N)\bar{C}(N)^T + V_2\right]\bar{K}(N)^T \end{cases}$$

Predicted $\delta\hat{x}(N+1|N) \neq x(N+1)$

by definition

$$\delta x(t) = x(t) - \bar{x}(t) \Rightarrow x(t) = \bar{x}(t) + \delta x(t)$$

$$\Downarrow$$

$$\hat{x}(N+1|N) = \bar{x}(N+1) + \underbrace{\delta\hat{x}(N+1|N)}_{\text{given by Kalman predictor}}$$

⚠ P.A.

> Kalman is predicting the state perturbance and no more the state.

Estimated state can be obtained by linearized kalman prediction LK

$$
LK : \begin{cases} \hat{x}(N+1|N) = f(N, \hat{x}(N|N-1)) + \bar{K}(N)e(N) \\ \hat{y}(N|N-1) = h(N, \hat{x}(N|N-1)) \\ e(N) = y(N) - \hat{y}(N|N-1) \end{cases}
$$

$\bar{K}(N)$ depends by $\bar{A}(N), \bar{C}(N)$ for the linearized system

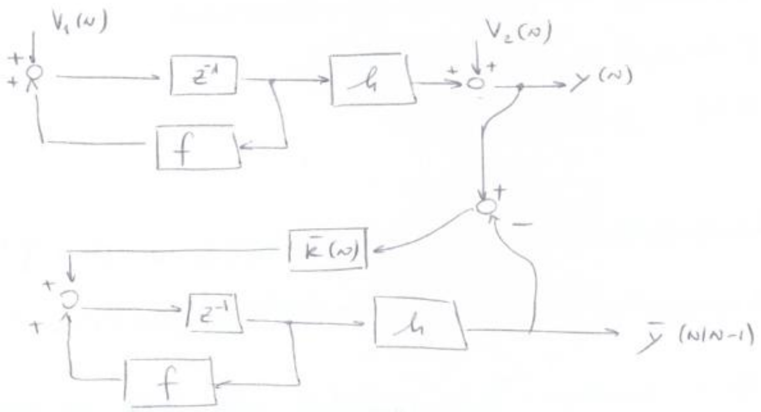
computed previously around nominal movement

$\Downarrow$

Real state is not taken in account

$\Downarrow$

Huge estimation error $\Rightarrow$ This strategy is not used at all



Extended KALMAN predictor is used in Robust Control System

In robust control

$$\begin{cases} x_1(t+1) = \underline{x_2(t)}\, x_1(t) + v_{11}(t) \\ \dots \\ x_2(t+1) = x_2(t) + v_{12}(t) \end{cases}$$

x_2 should be a gain changing

Non-linear

using extended kalman estimator

KALMAN THEORY LIMIT :

- Good solution strongly depends by assumptions (a priori information)

$\Downarrow$

Difficult to find out wich assumptions are correct or not