IDENTIFICATION TECHNIQUES

(SYSTEM)
IDENTIFICATION ≡ Estimation of dynamic models

Ⓐ Estimation ; Ⓑ dynamic models

Ⓐ **Estimation**
- → Parametric $(\theta \in ④)$
- → Non-parametric

EXAMPLE 1 $Y \sim p_\theta(Y) = \mathcal{N}(\theta, 1)$ θ = mean is the unknown parameter

$\underline{Y} := \{Y_1, \dots, Y_N\}$ Y_i i.i.d $\sim p_\theta(Y)$

$\downarrow$

$\underline{y} := \{y_1, \dots, y_N\}$ Sample values

Estimator: $\hat{\theta}(\underline{Y}) := T(Y_1, \dots, Y_N) = \frac{1}{N} \cdot \sum_{i=1}^{N} Y_i$

$\mathbb{E}\, \hat{\theta}(Y) = \theta$ (Unbiased)

$\text{Var}\, \hat{\theta}(Y) = \frac{1}{N}$

\+ $\underline{Y}$ GAUSSIAN

$\Downarrow$

$\hat{\theta}(\underline{Y}) \sim \mathcal{N}\left(\theta, \frac{1}{N}\right)$

EXAMPLE 2 (Non-parametric)

$Y \sim p(y)$ (Model) **DATA:** $\underline{Y} := \{Y_1, \dots, Y_N\}$

Estimator: $\hat{p}_Y(y) = \sum_{i=1}^{N} \alpha_i K(y, Y_i)$ → "Kernel"

e.g. $K_\sigma(y,x) = \exp\left\{-\frac{\|x-y\|^2}{\sigma^2}\right\}$ GAUSSIAN KERNEL

ORDER ESTIMATION

Criterions of the form (See BAUER, Automatica 2005)

$$IC(n) = J_n(\hat{S}) + \frac{d(n)\, C_N}{N}$$

$d(n) = 2mn + mp + np$

$m =$ # of outputs

$p =$ # of inputs

$$J_n(\hat{S}) = \begin{cases} 1.\ \sum_{i=n+1}^{M} \hat{\sigma}_i^{2} \\ 2.\ \hat{\sigma}_{n+1}^{2} \\ 3.\ \sum_{i=n+1}^{M} -\log\left(1 - \hat{\sigma}_i^{2}\right) \end{cases}$$

[FOR CANONICAL CORRELATION ANALYSIS]

Other Criterion (Innovation Variance Criterion)

$$IVC(n) = \log\det\left(\tilde{\Lambda}(n)\right) + \frac{d(n)\, C_N}{N}$$

$$\tilde{\Lambda}(n) = \frac{1}{N} \cdot \sum_{t=1}^{N} \underset{\uparrow}{\hat{e}_t(\hat{\theta}_n)}\, \hat{e}_t^{T}(\hat{\theta}_n)$$

Estimated Innovation using the parameter estimate $\hat{\theta}_n$ of order n

Otherwise:

Borrow Criterion from estimation of dimensionality in Canonical Correlation Analysis

FUJIKOSHI - VEITCH (Biometrika 1979) [for output only, slight changes with inputs!]

$$J(n) = -\sum_{i=n+1}^{N} \log\left(1-\hat{\sigma}_i^2\right) - \underset{\text{lenght of future}}{\underset{\uparrow}{(fm-n)}}\,\underset{\text{lenght of past.}}{\underset{\uparrow}{(pm-n)}}\,\frac{\cdot C_N}{N}$$

REVISITING SUBSPACE PROCEDURES

Consider model in PREDICTION FORM

$$X_{t+1} = (A - KC) X_t + K Y_t \qquad \bar{A} := A - KC$$

$$\nu := T - t + 1$$

$$Y_{[t,T]} = \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{\nu-1} \end{bmatrix} X_t + \begin{bmatrix} 0 & \ddots & & 0 \\ CK & \ddots & & \\ C\bar{A}K & \ddots & & \\ \vdots & \ddots & & CK \; 0 \end{bmatrix} Y_{[t,T]} + E_{[t,T]}$$

DEFINITION (OBLIQUE PROJECTION)

$\mathcal{A}, \mathcal{B} \subset \mathcal{H}$ (Hilbert space) $\mathcal{A} \cap \mathcal{B} = \{0\}$

$v \in \mathcal{H}$ $\qquad \hat{E}[v / \mathcal{A} \cup \mathcal{B}] = \hat{v}_A + \hat{v}_B$

$\hat{v}_A =: \hat{E}_{\parallel \mathcal{B}}[v | \mathcal{A}] \qquad \hat{v}_B =: \hat{E}_{\parallel \mathcal{A}}[v / \mathcal{B}]$

$\uparrow$ oblique projection Along $\mathcal{B}$ onto $\mathcal{A}$.

Computing oblique projections

$\mathcal{H} = \mathbb{R}^N$, A, B basis for $\mathcal{A}$ and $\mathcal{B}$

$$\hat{v}_A =: \hat{E}_{\parallel B}[v | A] = (v|B)(A|B)^T \left[(A|B)(A|B)^T \right]^{-1} A$$

$$A|B := A - AB^T (BB^T)^{-1} B$$

$$v|B := v - vB^T (BB^T)^{-1} B$$

Consider Infinite Past first

$$\hat{E}_{\| Y_{[t,t+k)}}\left[y(t+k) \,|\, Y_t^- \right] = C(A-KC)^k X_t =: \hat{Y}^{\text{obl}}_{t+k|t-1}$$

$$X_t = \sum_{i=1}^{+\infty} (A-KC)^{i-1} K Y_{t-i}$$

$$= \underbrace{(A-KC)^p X_{t-p}}_{} + \sum_{i=1}^{p} (A-KC)^{i-1} K Y_{t-i}$$

$$\left| \lambda_{\max}(A-KC) \right|^p \xrightarrow[p \to \infty]{} 0 \quad (\text{exponentially fast!})$$

$$\boxed{X_t \simeq \sum_{i=1}^{p} (A-KC)^{i-1} K Y_{t-i}}$$

DEFINE!

$$\hat{Y}^{\text{obl}}_{[t,T]} := \begin{bmatrix} \hat{Y}^{\text{obl}}_{t|t-1} \\ \vdots \\ \hat{Y}^{\text{obl}}_{T|t-1} \end{bmatrix} = \bar{O}_\nu X_t \simeq \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{\nu-1} \end{bmatrix} \left[\bar{A}^{p-1}K \;\; \bar{A}^{p-2}K \; \dots \; \bar{A}K \;\; K \right] \begin{bmatrix} Y_{t-p} \\ \vdots \\ Y_{t-1} \end{bmatrix}$$

$$\bar{O}_\nu := \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{\nu-1} \end{bmatrix}$$

From $\hat{Y}^{obl}_{[t,T]} = \bar{O}_\nu X_t$

1.) Choose W_p " weighting matrix "

2.) Do SVD

$$W_p^{-1} \hat{Y}^{obl}_{[t,T]} = USV \underset{\substack{\uparrow \\ \text{order estimation}}}{\simeq} U_n S_n V_n^T$$

3) $\hat{\bar{O}}_\nu := W_p U_n S_n^{1/2}$

4) <u>Estimate the state</u>

$$\hat{X}_t := S_n^{-1/2} U_n^T W_p^{-1} \hat{Y}^{obl}_{[t,T]}$$

$$\left(= \left(\hat{\bar{O}}_\nu^T W_p^{-T} W_p^{-1} \hat{\bar{O}}_\nu \right)^{-1} \hat{\bar{O}}_\nu^T W_p^{-T} W_p^{-1} \cdot \hat{Y}^{obl}_{[t,T]} \right)$$

$$\hat{X}_{t+1} := S_n^{-1/2} U_n^T W_p^{-1} \hat{Y}^{obl}_{[t+1,T+1]}$$

<u>Where:</u>

$$\hat{Y}^{obl}_{[t+1,T+1]} := \begin{bmatrix} \hat{Y}^{obl}_{t+1|t} \\ \vdots \\ \hat{Y}^{obl}_{T+1|t} \end{bmatrix}$$

5) Estimate

$$\hat{C} := Y_t \hat{X}_t^T (\hat{X}_t \hat{X}_t^T)^{-1}$$

$$\hat{E}_t := Y_t - C\hat{X}_t$$

$$\hat{A} := \hat{X}_{t+1} \hat{X}_t^T (\hat{X}_t \hat{X}_t^T)^{-1}$$

$$\hat{K} := \hat{X}_{t+1} \hat{E}_t (\hat{E}_t \hat{E}_t^T)^{-1}$$

$$\hat{\Lambda} := \frac{1}{N} \cdot \hat{E}_t \hat{E}_t^T$$

PROCEDURE WITH VARX IDENTIFICATION

1. Estimate $\hat{\Phi}_i$ solution of

$$Y_t \cong \sum_{i=1}^{p} \hat{\Phi}_i Y_{t-i}$$

in the Least squares sense

2. Compute

$$\hat{Y}^{obl}_{[t,T]} := \begin{bmatrix} \hat{\Phi}_p & \cdots & & \hat{\Phi}_1 \\ 0 & \hat{\Phi}_p & \cdots & \hat{\Phi}_2 \\ & \ddots & \ddots & \\ 0 & & 0 & \ddots \end{bmatrix} \begin{bmatrix} Y_{t-p} \\ \vdots \\ Y_{t-1} \end{bmatrix}$$

3. Follow Algorithm on page (70)

A NAIVE "SUBSPACE" APPROACH (AOKI)

1. Given Data $\{y(k)\}_{k=0,1,\dots,T}$

   Compute

$$\hat{\Lambda}(\tau) := \frac{1}{N} \sum_{i=0}^{N-\tau} y(i+\tau)\, y(i)^T$$

2. Form

$$\mathcal{H} := \begin{bmatrix} \hat{\Lambda}(1) & \hat{\Lambda}(2) \dots & \hat{\Lambda}(\nu) \\ \vdots & & \\ \hat{\Lambda}(\nu+1) \dots & & \hat{\Lambda}(2\nu) \end{bmatrix} \qquad \nu \text{ "big enough"}$$

3. Do Ho-Kalman on $\mathcal{H}$:

$$\mathcal{H} = USV^T \simeq U_n S_n V_n^T$$

$$S = \text{diag.}\{\sigma_1, \dots, \sigma_n, \sigma_{n+1} \dots \sigma_\nu\}$$

$$\sigma_{n+1} \simeq 0 \quad \dots \quad \sigma_\nu \simeq 0 \quad \left(\begin{matrix}\text{Can be} \\ \text{Neglected}\end{matrix}\right)$$

WHAT DOES "SMALL" MEAN?

This is a (statistical) model approximation step.

$$\hat{C} := U_n(1:m,:) \qquad \hat{\bar{C}}^T := V_n^T(:,1:m)$$

$$\hat{A} = \hat{O}_{\nu-1}^{+} \hat{\bar{O}}_{\nu} \qquad \text{whore} \qquad \hat{O}_{\nu} := \left[U_n S_n^{1/2} \right]$$

$$\hat{\bar{O}}_{\nu} := O_{\nu}(m+1:\text{end},:)$$

$$\hat{O}_{\nu-1} := \left[\hat{O}_{\nu}(1:\text{end}-m,:) \right]$$

4) GIVEN $\hat{A}, \hat{C}, \hat{\bar{C}}, \frac{1}{2}\hat{\Lambda}(0)$ find the minimal solution P_- of

$$P - \hat{A}P\hat{A}^T - (\hat{\bar{C}}^T - \hat{A}P\hat{C}^T)(\hat{\Lambda}(0) - \hat{C}P\hat{C}^T)^{-1}(\hat{\bar{C}}^T - \hat{A}P\hat{C}^T)^T = 0$$

Then The estimated model is:

$$x(t+1) = \hat{A}x(t) + \hat{K}e(t) \qquad \hat{\Lambda}_e = \text{Var}\{e(t)\}$$
$$y(t) = \hat{C}x(t) + e(t)$$

where
$$\begin{cases} \hat{K} := (\hat{\bar{C}}^T - \hat{A}P_-\hat{C}^T)(\hat{\Lambda}(0) - \hat{C}P_-\hat{C}^T)^{-1} & \text{(Kalman Gain)} \\ \hat{\Lambda}_e := (\hat{\Lambda}(0) - \hat{C}P_-\hat{C}^T) & \text{(Innovation Variance)} \end{cases}$$

SUBSPACE IDENTIFICATION

BASIC IDEA: $y(t)$ is measured, if $x(t)$ and $x(t+1)$ where available we could solve

$$\begin{aligned} x(t+1) &= A\,x(t) + K\,e(t) \\ y(t) &= C\,x(t) + e(t) \end{aligned}$$

[$x(t)$ State of the (forward) Kalman filter]

in the Least squares sense.

TO DO SO WE NEED TO COMPUTE THE COVARIANCES

$$\mathbb{E}\; x(t)x^T(t), \qquad \mathbb{E}\; x(t+1)x^T(t) \qquad \mathbb{E}\; y(t)\,x^T(t)$$

(can use sample version!)

$$A = \Big[\mathbb{E}\; x(t+1)\,x^T(t)\Big]\Big[\mathbb{E}\; x(t)\,x^T(t)\Big]^{-1} \simeq \Big[\frac{1}{N}\cdot\sum_{i=1}^{N-1} x(1+i)\,x^T(i)\Big]\Big[\frac{1}{N}\cdot\sum_{i=1}^{N} x(i)\,x^T(i)\Big]^{-1}$$

$$C = \Big[\mathbb{E}\; y(t)\,x^T(t)\Big]\Big[\mathbb{E}\; x(t)\,x^T(t)\Big]^{-1} \simeq \Big[\frac{1}{N}\cdot\sum_{i=1}^{N} y(+i)\,x^T(i)\Big]\Big[\frac{1}{N}\cdot\sum_{i=1}^{N} x(i)\,x^T(i)\Big]^{-1}$$

what about $\bar{C}^T$?

Note that $\bar{C}^T = APC^T + BD^T$

$$\overset{!}{=} \mathbb{E}\Big[\, x(t+1)\, y^T(t)\Big]$$

$$\bar{C}^T = \mathbb{E}\Big[x(t+1)\,y^T(t)\Big] \simeq \frac{1}{N}\cdot\sum_{i=1}^{N-1} x(1+i)\,y^T(i).$$

NEED PROCEDURES TO CONSTRUCT A BASIS FOR THE STATE SPACE!

Preliminaries: HILBERT SPACE OF A STATIONARY SIGNAL

$\{y(t)\}_{t\in\mathbb{Z}}$ stationary (zero mean and finite variance) stochastic process

$$\mathcal{H} := \{\eta : \eta = \sum_i a_i^T y(t_i) \quad a_i \in \mathbb{R}^m\}$$

with inner product $\langle \eta, \xi \rangle := \mathbb{E}[\eta \xi]$

- "Interesting" Subspaces

PAST: $\mathcal{H}_t^- := \{\eta : \eta = \sum_i a_i^T y(t_i) \quad a_i \in \mathbb{R}^m, \ t_i < t\}$

FUTURE $\mathcal{H}_t^+ := \{\eta : \eta = \sum_i a_i^T y(t_i) \quad a_i \in \mathbb{R}^m, \ t_i \geq t\}$

$$\mathcal{H}_{[\bar{t}_m, \bar{t}_M]} := \{\eta = \sum_i a_i^T y(t_i), \ a_i \in \mathbb{R}^m \quad \bar{t}_m \leq t \leq \bar{t}_M\}$$

- Let $x(t)$ be the state vector of the (steady state) Kalman filter

$$x(t+1) = A x(t) + K e(t)$$
$$y(t) = C x(t) + e(t)$$

$$\mathcal{X}_t := \text{span}\{x(t)\} = \{a^T x(t), \ a \in \mathbb{R}^n\}$$

REMARK:

$$x(t) = \sum_{i=1}^{+\infty} K(A-KC)^{i-1} y(t-i) \in H_t^-$$

$$\forall\, k \quad y(t+k) = CA^{k-1}x(t) + \sum_{i=1}^{k-1} CA^{i-1}K\, e(t+k-i) + e(t)$$

- Let $\hat{E}[\cdot | \mathcal{A}]$ be the orthogonal projection (in $\mathcal{H}$) onto the subspace $\mathcal{A} \subseteq \mathcal{H}$
- $e(t)$ (the innovation process) is uncorrelated with $y(s)$ $\forall\, s < t$ ($e(t)$ is by definition, given by $e(t) := y(t) - \hat{E}[y(t) / H_t^-]$, which is orthogonal to H_t^-)

$$\Rightarrow \quad \hat{E}[y(t+k) | H_t^-] = CA^k x(t)$$

$$\downarrow$$

$$\boxed{\hat{E}[H_t^+ | H_t^-] = X_t} \qquad (*)$$

$\begin{Bmatrix} x(t) \\ y(t) \end{Bmatrix}$ is

HOMEWORK: $\mathcal{X}(t)$ is Markovian i.e.

$$\mathbb{E}\left[x(t+1) \mid \mathcal{X}_{t+1}^{-}\right] = \mathbb{E}\left[x(t+1) \mid \mathcal{X}_t\right] \quad \left(= A x(t)\right)$$

Relation (*) SUGGESTS A GEOMETRIC PROCEDURE TO CONSTRUCT THE STATE SPACE.

Remark: for finite dimensional models ($\dim x(t) < +\infty$)

$$\hat{\mathbf{E}}\left[\mathcal{H}_{[t,t+k)} \mid \mathcal{H}_t^{-}\right] = \mathcal{X}_t \quad \forall \; k \geq n$$

i.e need only a finite number of projections.

IDEAL ALGORITHM:

1. compute $\hat{y}(t+k|t-1) := \hat{E}\left[y(t+k) / \mathcal{H}_t^{-}\right]$

$$k = 0, 1, \ldots f \geq n+1$$

$$\mathcal{X}_t = \operatorname{span}\left\{\hat{y}(t|t-1), \ldots, \hat{y}(t+f-1|t-1)\right\}$$

$$\mathcal{X}_{t+1} = \operatorname{span}\left\{\hat{y}(t+1|t), \ldots, \hat{y}(t+f|t)\right\}$$

2. Choose basis (coherent) $x(t)$, $x(t+1)$

3. Compute $A, C, \bar{C}$, as on page 55

PRACTICAL "PROBLEMS"

1. How do we translate these "abstract" procedures into algorithms working with data?

2. Infinite past ($\mathcal{H}_t^-$) is never available!

"SOLUTIONS"

1. Isometric Isomorphism between random variables $y(t)$ and data sequences

$$Y_t := [\, y_t, y_{t+1}, \dots, y_{t+N-1}, \dots \,]$$

$\uparrow$ MATRIX FORMED WITH DATA.

DETAILS LATER ON

2. USE FINITE PAST ($\mathcal{H}_{[t-p,t)}$) and use the TRANSIENT Kalman filter.

DETAILS LATER ON.

FROM RANDOM VARIABLES TO DATA

$$y(t) \xleftrightarrow{1:1} [\, y_t \; y_{t+1} \; y_{t+2} \cdots \; y_{t+N-1} \cdots \,] = Y_t^{\infty}$$

(Infinitely long tail)

$(y_t \equiv$ sample value of $y(t))$

Inner Product

$$\mathbb{E}\, y(t)\, y^T(s) \longleftrightarrow \lim_{N\to\infty} \frac{1}{N} \cdot \sum_{i=0}^{N-1} y_{t+i}\, y^T_{s+i}$$

$$\left(= \langle y(t), y(s) \rangle \right) \qquad \left(= \langle Y_t^{\infty}, Y_s^{\infty} \rangle \right)$$

$$\mathcal{H}_{[t,s]} = \text{span}\left\{ y(\tau),\ \tau \in [t,s] \right\} \longleftrightarrow \text{row space of} \begin{bmatrix} Y_t^{\infty} \\ Y_{t+1}^{\infty} \\ \vdots \\ Y_s^{\infty} \end{bmatrix} =: Y_{[t,s]}^{\infty}$$

IN PRACTICE! Finite Data, consider only FINITE tails

$$Y_t := [\, y_t, y_{t+1}, \dots, y_{t+N-1} \,]$$

SUBSPACE .. IDENTIFICATION

Repeat the idel procedure described with random variables using (Finite) tails of data (inner product is the standard inner product between vectors in $\mathbb{R}^N$)

PAST DATA $Y_{[t_0,t)} := \begin{bmatrix} Y_{t_0} \\ Y_{t_0+1} \\ \vdots \\ Y_{t-1} \end{bmatrix}$

Block Hankel data matrices

FUTURE DATA $Y_{[t,T]} := \begin{bmatrix} Y_t \\ \vdots \\ Y_T \end{bmatrix}.$

STATE : $X_t := [x_t \; x_{t+1} \dots \; x_{t+N-1}]$

Relation with block Hankel data matrices :

$$Y_{[t,T]} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{T-t} \end{bmatrix}}_{O_\nu} X_t + H_\nu^s E_{[t,T]}$$

Innovation.

where:

$$H_\nu^s := \begin{bmatrix} I & 0 & & & 0 \\ CK & I & \ddots & & \\ CAK & \ddots & \ddots & & \\ & \ddots & \ddots & & 0 \\ & & CAK & CK & I \end{bmatrix} \qquad E_{[t,T]} = \begin{bmatrix} e_t & e_{t+1} & \dots & e_{t+N-1} \\ e_{t+1} & e_{t+2} & \dots & e_{t+N} \\ & & & \\ e_T & e_{T+1} & \dots & e_{T+N-1} \end{bmatrix}$$

with the property that, for $N \to \infty$ the rows of $E_{[t,T]}$ are orthogonal to the rows of $Y_{[t_0,t)}$

BASIC STEPS:

1. Compute $\hat{Y}_{[t,T]} := \hat{E}[Y_{[t,T]} \mid Y_{[t_0,t)}]$

$$\simeq O_\nu \hat{X}_t$$

$$\uparrow$$

$$\frac{E_{[t,T]} \cdot Y^T_{[t_0,t)}}{N} \simeq 0$$

2. From $\hat{Y}_{[t,T]}$ find a basis $\hat{X}_t$

Note that

$$\text{row span } \hat{Y}_{[t,T]} = \text{row span } \hat{X}_t$$

Use Canonical correlation analysis (CCA)

Find the first n principal direction in $Y_{[t_0,t)}$ w.r.t. $Y_{[t,T]}$

i.e.

a) $$\underbrace{\left(Y_{[t,T]} Y^T_{[t,T]}\right)^{-1/2}}_{\hat{\Sigma}_{ff}^{-1/2}} \underbrace{\left(Y_{[t,T]} Y^T_{[t_0,t)}\right)}_{\hat{\Sigma}_{fp}} \underbrace{\left(Y_{[t_0,t)} Y^T_{[t_0,t)}\right)^{-T/2}}_{\hat{\Sigma}_{pp}^{-T/2}}$$

p = past f = future

$$= \hat{U}\hat{S}\hat{V}^T \underset{\uparrow}{\simeq} \hat{U}_n \hat{S}_n \hat{V}_n$$

order estimation step.

for $N \longmapsto \infty$

$$\hat{S} \longrightarrow S = \operatorname{diag}\{\sigma_1, \dots, \sigma_n, 0, \dots 0\}$$

$\sigma_i = \cos(\theta_i)$

- θ_i = principal angles between $\mathcal{H}_{[t,T]}$ and $\mathcal{H}_{[t_0,t)}$.
- n = dimension of the "true" system generating the data

b) Estimate of the observability matrix

$$\hat{O}_\nu := \hat{\Sigma}_{ff}^{+1/2} \hat{U}_n \hat{S}_n^{1/2}$$

Estimate of the state

$$\hat{X}_t := \hat{S}_n^{1/2} \hat{V}_n^T \hat{\Sigma}_{pp}^{-1/2} Y_{[t_0,t)}$$

Estimate of the state at time t+1

$$\hat{X}_{t+1} := \hat{S}_n^{-1/2} \, \hat{U}_m^T \, \hat{\Sigma}_{ff}^{-1/2} \cdot \hat{Y}_{[t+1,T+1]}$$

where
$$\hat{Y}_{[t+1,T+1]} := E\left[Y_{[t+1,T+1]} \mid Y_{[t_0,t+1)} \right]$$

Remark : this slightly differs from the procedure described in, e.g., Van Overschee - De Moor Automatica 1993.

c) Having the state :

SOLVE
$$\hat{X}_{t+1} = A\hat{X}_t + K_t \hat{E}_t$$
$$Y_t = C\hat{X}_t + \hat{E}_t$$
in the least squares sense for A, C

$$\hat{A} := \left(\hat{X}_{t+1} \hat{X}_t^T \right) \left(\hat{X}_t \hat{X}_t^T \right)^{-1}$$
$$\hat{C} := \left(Y_t \hat{X}_t^T \right) \left(\hat{X}_t \hat{X}_t^T \right)^{-1}$$

Estimate $\bar{C}^T$ as :

$$\hat{\bar{C}}^T := \frac{\left(\hat{X}_{t+1} Y_t^T \right)}{N}$$

Estimate $\hat{\Lambda}(0) := \frac{Y_t Y_t^T}{N}$

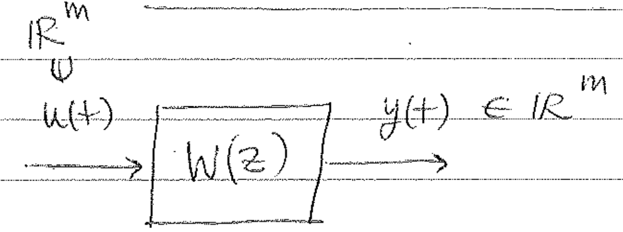
d) Given $\left(\hat{A}, \hat{C}, \hat{\bar{C}}^T, \hat{\Lambda}(0)\right)$

find $\hat{P}_-$ solution of LMI

$$\hat{K} := \left(\hat{\bar{C}}^T - \hat{A}\hat{P}_-\hat{C}^T\right)\hat{\Lambda}_e^{-1} \qquad \text{Kalman Gain}$$

$$\hat{\Lambda}_e := \left(\hat{\Lambda}(0) - \hat{C}\hat{P}_-\hat{C}^T\right) \qquad \text{Innovation Variance.}$$

STATE SPACE MODELS OF DETERMINISTIC SYSTEMS



$$\begin{cases} x(t+1) = Ax(t) + Bu(t) & x(t_0) = x_0 \qquad t \geq t_0 \\ y(t) = Cx(t) + Du(t) \end{cases}$$

$\{x(t)\}$ state $\in \mathbb{R}^n$

TRANSFER FUNCTION $W(z) = C(zI - A)^{-1}B + D$

$$= C \frac{\operatorname{adj}(zI - A)}{\det(zI - A)} B + D$$

OBSERVABILITY $\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}$ rank $\mathcal{O} = n$

REACHABILITY $\mathcal{R} = [B \ AB \dots A^{n-1}B]$ rank $\mathcal{R} = n$

Thm:

MINIMALITY (i.e. $x(t)$ "as small as possible")

$\Updownarrow$

OBSERVABLE + REACHABLE

(minimal systems)

p_k = poles $W(z)$ = roots of $\left[\det(zI - A)\right]$ = eig(A)

BIBO STABILITY $\longleftrightarrow$ $|\text{eig}(A)| < 1$

ZEROS ? if $m = p$, D invertible

can define "inverse system"

$$\begin{cases} x(t+1) = (A - BD^{-1}C)x(t) + BD^{-1}y(t) \\ u(t) = -D^{-1}Cx(t) + D^{-1}y(t) \end{cases}$$

$$\text{zeros } W(z) = \text{poles}\left(W^{-1}(z)\right) = \text{eig}\left(A - BD^{-1}C\right)$$

MORE IN GENERAL :

z_0 is a zero (Transmission)

if $\begin{bmatrix} -(z_0 I - A) & B \\ C & D \end{bmatrix}$ is rank deficent (Assume full rank generically

(for $m \geq p$)

FIND $\begin{bmatrix} \theta \\ \eta \end{bmatrix}$: $\begin{bmatrix} -(z_0 I - A) & B \\ C & D \end{bmatrix} \begin{bmatrix} \theta \\ \eta \end{bmatrix} = 0$

Homework: $(m \geq p)$

· SHOW THAT if z_0 is a zero with (right) nulling vector $\begin{bmatrix}\theta \\ \eta\end{bmatrix}$, then

if $x(0) = \theta$, $u(t) = \eta (z_0)^t$ it follows

that $y(t) \equiv 0$ (and $x(t) = \theta z_0^t$) $t \geq 0$

Homework.

SHOW THAT if D is invertibe

Transmission zeros $=$ eig. $(A - BD^{-1}C)$

SINGULAR VALUE DECOMPOSITION

$A \in \mathbb{R}^{m \times p}$ rank $A = n$ $\left(\leq \min(m,p) \right)$

$\exists\ U \ \left(\in \mathbb{R}^{m \times m} \right) \quad U^T U = U U^T = I$

$\exists\ V \ \left(\in \mathbb{R}^{p \times p} \right) \quad V^T V = V V^T = I$

s.t.

$$A = U S V^T \qquad S = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Sigma = \text{diag.} \{ \sigma_1, \dots, \sigma_n \} \qquad \sigma_1 \geq \sigma_2 \geq \sigma_3 \dots \geq \sigma_n$$

$\uparrow$ $n =$ rank A.

$$A = U_n \Sigma V_n^T \qquad U_n = [u_1 \dots u_n] \qquad U_n^T U_n = I.$$

$$V_n = [v_1 \dots v_n] \qquad V_n^T V_n = I$$

$$= \sum_{i=1}^{n} u_i v_i^T \cdot \sigma_i = \sum_{i=1}^{n} \underset{\substack{\uparrow \\ [m,p] \text{ rank } 1}}{A_i} \sigma_i$$

INTERPRETATION: (*)

$$\sigma_1 = \sup_{\substack{v_1 \\ \|v_1\| = 1}} \| A v_1 \|_2$$

$$\sigma_2 = \sup_{\substack{v_2 \perp v_1 \\ \| v_2 \| = 1}} \| A v_2 \|_2$$

$\vdots$

etc.

FACT: Using the construction procedure in (*)

$V^T V = I$ $\qquad U^T U = I$

(a) $\qquad$ (b)

Pf: (a) by construction.

Let $\hat{U} = \left[u_1 \;\; U_1^{\perp} \right]$ $\qquad \hat{V} = \left[v_1 \;\; V_1^{\perp} \right]$

$$\hat{U}^T A \hat{V} = \begin{bmatrix} u_1^T \\ (U_1^{\perp})^T \end{bmatrix} A \left[v_1 \;\; V_1^{\perp} \right] = \begin{bmatrix} \sigma_1 & u_1^T A V_1^{\perp} \\ (U_1^{\perp})^T A v_1 & (U_1^{\perp})^T A V_1^{\perp} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1 & u_1^T A V_1^{\perp} \\ 0 & A_1 \end{bmatrix} \qquad A_1 := (U_1^{\perp})^T A V_1^{\perp}$$

Show that $u_1^T A V_1^{\perp} = 0$ $\qquad$ Let $v \in V_1^{\perp}$, $\|v\| = 1$

and let $\bar{v} := \frac{1}{\sigma_1} v_1 + \alpha \frac{\sigma_{12}}{\sigma} v$ $\qquad \sigma = \| A v \|_2$

$$\bar{u} := \frac{A \bar{v}}{\| A \bar{v} \|} = \frac{u_1 + \alpha \sigma_{12} u_2}{\| u_1 + \sigma_{12} u_2 \|}$$

$$\| u_1 + \sigma_{12} u_2 \| = 1 + \alpha^2 \sigma_{12}^2 + 2 \sigma_{12}^2 \alpha =$$

$$\| \bar{v} \| = \sqrt{\frac{1}{\sigma_1^2} + \frac{\alpha^2 \sigma_{12}^2}{\sigma^2}} = \sqrt{\frac{\sigma^2 + \sigma_{12}^2 \sigma_1^2 \alpha^2}{\sigma_1^2 \sigma^2}}$$

$$\frac{\| A \bar{v} \|}{\| \bar{v} \|} = \sigma_1 \sqrt{\frac{\sigma^2}{\sigma^2 + \alpha^2 \sigma_{12}^2 \sigma_1^2}} \left(1 + \alpha^2 \sigma_{12}^2 + 2 \sigma_{12}^2 \alpha \right)$$

$$\frac{\|A\bar{v}\|}{\|\bar{v}\|} = \sigma_1 \cdot \frac{1 + \alpha^2 \sigma_{12}^2 + 2\sigma_{12}^2 \alpha}{\sqrt{1 + \alpha^2 \frac{\sigma_{12}^2 \sigma_1^2}{\sigma^2}}}$$

$$\left.\frac{\partial \frac{\|A\bar{v}\|}{\|\bar{v}\|}}{\partial \alpha}\right|_{\alpha=0} = \frac{2\sigma_{12}^2}{1} = 2\sigma_{12}^2 > 0$$

$\Rightarrow \exists\, \alpha > 0 : \frac{\|A\bar{v}\|}{\|\bar{v}\|} > \sigma_1$ Assurdo contro l'Ip che σ_1 è il max $\max_{v} \frac{\|Av\|}{\|v\|}$

$$\Rightarrow \hat{U}^T A \hat{V} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & A_1 \end{bmatrix}$$

SVD DECOMPOSE A_1 in $(U_1^\perp, V_1^\perp)$

FACT!

U = eigenvectors of AA^T

$[u_1, \dots, u_n]$

V = eigenvectors of A^TA

S^2 = (non zero) eigenvalues of A^TA and AA^T

MATRIX NORMS

$A \in \mathbb{R}^{m\times p}$

(TWO - NORM)

$$\|A\|_2 = \sup_{v\in\mathbb{R}^p} \frac{\|Av\|_2}{\|v\|_2} = \sigma_1$$

$$\|A_F\| = \sqrt{\sum_{i,j} a_{ij}^2} = \text{Tr.}\{A^TA\} = \text{Tr.}\{AA^T\}$$

$$= \sum_{i=1}^{n} \sigma_i^2$$

(FROBENIUS NORM)

LEAST SQUARES PROBLEMS AND CONDITION NUMBER.

$y = Ax$ let $\underline{y} = y + \Delta y$ How does $\bar{x}$

$: x\ \underline{y} = A\underline{x}$ changes?

$\underline{x} = x + \Delta x$

Compute $\dfrac{\|\Delta x\| / \|x\|}{\|\Delta y\| / \|y\|}$

$\dfrac{\|\Delta x\|}{\|x\|} \rightarrow$ Relative perturbation on the solution

$\dfrac{\|\Delta y\|}{\|y\|} \rightarrow$ Relative perturbation on the "DATA"

$$C(A) := \sup_{y,\Delta y} \frac{\|\Delta x\|}{\|x\|}\cdot\frac{\|y\|}{\|\Delta y\|} = \sup_{y,\Delta y} \overbrace{\frac{\|y\|}{\|x\|}}^{\text{Large}} \cdot \overbrace{\frac{\|\Delta x\|}{\|\Delta y\|}}^{\text{small}}$$

$$\sup_{y} \frac{\|y\|}{\|x\|} = \sigma_{max}(A) \qquad \inf_{\Delta y} \frac{\|\Delta y\|}{\|\Delta x\|} = \sigma_{min}(A)$$

$$\Rightarrow \boxed{C(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)} \qquad \text{CONDITION NUMBER}}$$

(ALSO: $C(A) = \|A\|_2 \cdot \|A^+\|_2$)

$$\left(\begin{aligned} \|A\|_2 &= \sigma_1 = \sigma_{max}(A) \\ \|A^+\|_2 &= \sigma_{max}\left(A^+\right) = \frac{1}{\sigma_{min}(A)} \end{aligned} \right)$$

$A^{\dagger}$ = MOORE-PENROSE PSEUDOINVERSE

$$= U_n S_n^{-1} V_n^T = \sum_{i=1}^{n} u_i v_i^T \cdot \frac{1}{\sigma_i}$$

$$= \underbrace{\bar{U}_n \bar{\Sigma}_n \bar{V}_n^T}_{\uparrow}$$

<u>Riordinati</u> (SVD of $A^{\dagger}$)

ANGLES BETWEEN SUBSPACES AND SVD

$\mathcal{A}$ = row span $\{A\}$ $\quad\mathcal{B}$ = row span $\{B\}$ $\quad\mathcal{A}, \mathcal{B} \subset \mathbb{R}^n$

$A \in \mathbb{R}^{p\times n}$ $\quad B \in \mathbb{R}^{m\times n}$

$p < n$ $\quad m < n$

(1.) GEOMETRIC PROCEDURE

(a) find a_1, b_1 $\quad a_1^T \in \mathcal{A}$ $\quad b_1^T \in \mathcal{B}$

$\|a_1\| = \|b_1\| = 1$ and $\sigma_1 := a_1^T b_1$ be maximal

a_1, b_1 first PRINCIPAL DIRECTIONS

$\theta_1 = a\cos\sigma_1$ first (minimal) angle between $\mathcal{A}$ and $\mathcal{B}$

(b) repeat iteratively:

(i) $\mathcal{A}_j := \text{span}\{a_1, \ldots, a_j\}$ $\quad\mathcal{B}_j := \text{span}\{b_1, \ldots, b_j\}$

$\mathcal{A}_j^\perp := \{$ orth. compl. of $\mathcal{A}_j$ in $\mathcal{A}\}$

$\mathcal{B}_j^\perp := \{$ orth. compl. of $\mathcal{B}_j$ in $\mathcal{B}\}$

(ii) Compute a_{j+1}, b_{j+1}, $\sigma_{j+1} = \cos(\theta_{j+1})$ as principal directions and minimal angle between $\mathcal{A}_j^\perp$ and $\mathcal{B}_j^\perp$.

ANGLES VIA SVD

$$\bar{A} := (AA^T)^{-1/2} A \qquad (\bar{A}\bar{A}^T = I)$$

$$\bar{B} := (BB^T)^{-1/2} B \qquad (\bar{B}\bar{B}^T = I)$$

$$\Sigma_{AB} := \bar{A}\bar{B}^T = (AA^T)^{-1/2} AB^T (BB^T)^{-T/2}$$

Take SVD:

$$USV = \Sigma_{AB}$$

DEFINE $\hat{A} := U^T\bar{A} = U^T(AA^T)^{-1/2}A \qquad (\hat{A}\hat{A}^T = I)$

$$\hat{B} := V^T\bar{B} = V^T(BB^T)^{-1/2}B \qquad (\hat{B}\hat{B}^T = I)$$

$$\hat{A}\hat{B}^T = U^T\bar{A}\bar{B}^T V = S = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$$

$\hat{A}$ = Basis for $\mathcal{A}$ $\{\text{span}\{\hat{A}\} = \text{span}\{A\}\}$

$\hat{B}$ = Basis for $\mathcal{B}$

$$\sup_{\substack{u,v \\ (\|\hat{A}^T u\| = \|\hat{B}^T v\| = 1)}} (u^T\hat{A})(\hat{B}^T v) = \sup_{\substack{u,v \\ \|u\| = \|v\| = 1}} u^T S v \Rightarrow u = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad v = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\sup_{\|u\| = \|v\| = 1} u^T S v = \sigma_1 = \cos \underset{\substack{\uparrow \\ \text{minimal angle}}}{\theta_1}$$

SVD and Reduced Rank matrix approximation

Let $A \in \mathbb{R}^{p \times m}$

find $\hat{A}_m$ "best" approximation of A of

fixed rank n $(\leq m)$

$$\hat{A}_m := \underset{\substack{A_n \in \mathbb{R}^{p \times m} \\ \text{rank } A_m \leq m}}{\arg\min} \| A - A_n \|_2^2 = \underset{\substack{A_n \in \mathbb{R}^{p \times m} \\ \text{rank } A_n \leq n}}{\arg\min} \| A - A_n \|_F^2$$

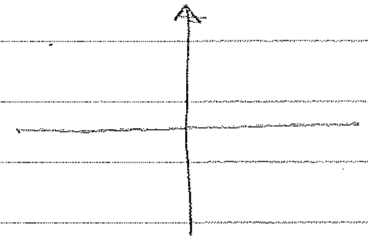
FACT: $\hat{A}_m = U_n S_n V_n^T$ where $A = USV$

and ~~A~~ $U_n = U_{1:n,:}$ $V_n = V_{1:n,:}$ $S_n = S_{1:n,1:n}$

TRANSFER FUNCTIONS

- $\mathcal{Z}$-Transform (ZETA)

$$f(t) \xmapsto{\mathcal{Z}} \begin{matrix}\hat{f}(z) = \\ F(z) = \end{matrix} \Big] = \sum_{t=-\infty}^{+\infty} f(t) z^{-t} \qquad z \in \mathbb{C}$$

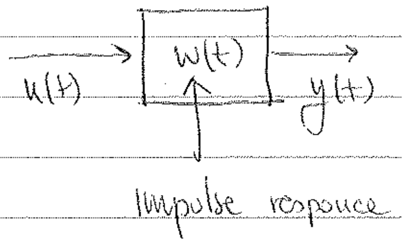


- $\mathcal{F}$-Transform (FOURIER)

$$\begin{matrix}\hat{f}(e^{j\theta}) = \\ F(e^{j\theta}) = \end{matrix} \Big] = \sum_{t=-\infty}^{+\infty} f(t) e^{-j\theta t}$$

STATE SPACE

$$\begin{aligned} X(t+1) &= A x(t) + B u(t) \\ y(t) &= C x(t) + D u(t) \end{aligned} \xrightarrow{\mathcal{Z}} W(z) = C(zI-A)^{-1}B + D$$



$$Y(z) = W(z)U(z)$$

$$y(t) = [W * u](t)$$

SIMILARITY TRANSFORMATION

$T \in \mathbb{R}^{n \times n}$ invertible

$\bar{x}(t) := T^{-1} x(t)$

$\bar{A} = T^{-1} A T \quad \bar{B} = T^{-1} B \quad \bar{C} = C T \quad \bar{D} = D$

$\bar{W}(z) = W(z)$

BALANCING (Stable Systems)

Ω Observability Grammian

$$\Omega = \sum_{k=0}^{+\infty} (A^T)^k C^T C A^k \qquad \Omega = A^T \Omega A + C^T C$$

Π Reachability Grammian

$$\Pi = \sum_{k=0}^{+\infty} A^k B B^T (A^T)^k \qquad \Pi = A \Pi A^T + B B^T$$

LYAPUNOV EQUATION:

$$P = A P A^T + B B^T \quad (*)$$

Any Two of the conditions below imply the third.

i) $|\lambda(A)| < 1$

ii) (A, B) Reach

iii) $\exists!$ solution $P = P^T > 0$ to $(*)$

INTERPRETATION (Reachability)

$$\Pi = \sum_{k=0}^{+\infty} A^k B B^T (A^T)^k = \mathcal{R}_\infty \mathcal{R}_\infty^T \qquad \mathcal{R}_\infty := [B \; AB \; A^2B \; \dots\dots]$$

Let $u(k)$: $\sum_{k=-\infty}^{-1} u^T(k)\, u(k) = 1$ $\quad (\| u \|_2 = 1)$

$$x(0) = \sum_{k=-1}^{-\infty} A^{k-1} B\, u(k) = \mathcal{R}_\infty \cdot \underline{u}$$

$$\langle x(0), x(0) \rangle = \langle \mathcal{R}_\infty \underline{u}, \mathcal{R}_\infty \underline{u} \rangle = \langle \underline{u}, \mathcal{R}_\infty^* \mathcal{R}_\infty \underline{u} \rangle$$

$$\sup_{\substack{\underline{u} \\ \|\underline{u}\|_2 = 1}} \underline{\langle x(0), x(0) \rangle} = \sigma_{max}(\mathcal{R}_\infty^* \mathcal{R}_\infty) = \sigma_{max}(\mathcal{R}_\infty \mathcal{R}_\infty^*)$$

$$= \sigma_{max}(\Pi)$$

$$\Pi = U_c \Lambda_c U_c^T$$

Change basis:

$$X_c(t) := U_c^T x(t)$$

$$B_c = U_c^T B \qquad A_c = U_c^T A U_c$$

$$\Rightarrow \mathcal{R}_c = U_c^T \cdot \mathcal{R}$$

$$\Omega_c = \Lambda_c \text{ (Diagonal) } = \text{diag}\{\underset{\substack{\uparrow \\ \text{Energy gains}}}{\lambda_{1,c}}, \dots, \lambda_{n,c}\}$$

For a unit energy input $\nearrow$ energy λ_c along first comp. of x_c

$\searrow$ energy $\lambda_{n,c}$ along last comp. of x_c $\uparrow$

certain state component nearly unexcited.

INTERPRETATION (OBSERVABILITY)

$$\Omega = \sum_{k=0}^{+\infty} (A^T)^k C^T C A^k \qquad \mathcal{O}_\infty = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ \vdots \end{bmatrix}$$

$$\Omega = \mathcal{O}_\infty^T \mathcal{O}_\infty$$

Let $x(0)$ s.t. $\|x(0)\|_2 = 1$

$$\underline{y} := \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(n) \\ \vdots \end{bmatrix} = \mathcal{O}_\infty x(0)$$

$$\langle \underline{y}, \underline{y} \rangle = \langle \mathcal{O}_\infty x(0), \mathcal{O}_\infty x(0) \rangle = \langle \mathcal{O}_\infty^* \mathcal{O}_\infty x(0), x(0) \rangle$$

$$\sup_{\substack{x(0) \\ \|x(0)\|=1}} \langle \mathcal{O}_\infty^* \mathcal{O}_\infty x(0), x(0) \rangle = \sigma_{max}(\mathcal{O}_\infty^* \mathcal{O}_\infty) = \sigma_{max}(\Omega)$$

$$\Omega = V_o \Lambda_o V_o^T \qquad \Lambda_c = \{\lambda_{1,o}, \dots, \lambda_{n,o}\}$$

$\uparrow$

state to output energy gains

If $\lambda_{1,o} \gg \lambda_{n,o}$

VERY VISIBLE ($x_{o,1}$)

NEARLY INVISIBLE ($x_{o,n}$)

$x_o(t) := V_o\, x(t)$ (in this special basis)

BALANCING: Choose (if possible!) state basis s.t. state components have the same input to state and state to outputs energy gains.

(i.e. $\overline{\Omega}$ and $\overline{\Pi}$ are DIAGONAL AND EQUAL)

PROCEDURE

1) $\Omega = U_0 S_0 U_0^T$

$\bar{x}(t) = T_1^{-1} x(t)$ $\quad T_1^T = S_0^{-1/2} U_0^T \iff T_1^{-1} = S_0^{1/2} U_0^T$

$\left(\text{s.t. } \bar{\Omega} = T_1^T \Omega T_1 \right) \quad \bar{\Omega} = I$

2) Compute SVD of $\bar{\Pi} := T_1^{-1} \Pi T_1^{-T}$

$= S_0^{1/2} U_0^T \Pi U_0 S_0^{1/2}$

$= V_0 \Lambda_0^2 V_0^T$

$T_2^{-1} = \Lambda_0^{-1/2} V_0^T$

s.t.

$\bar{\Pi}_2 = T_2^{-1} \bar{\Pi} T_2^{-T} = \Lambda_0^{-1/2} V_0^T V_0 \Lambda_0^2 V_0^T V_0 \Lambda_0^{-1/2} = \Lambda_0$

$\bar{\Omega}_2 = T_2^T \bar{\Omega} T_2 = \Lambda_0^{1/2} V_0^T V_0 \Lambda_0^{+1/2} = \Lambda_0$

(3) $T^{-1} = T_2^{-1} T_1^{-1} = \Lambda_0^{-1/2} V_0^T S_0^{1/2} U_0^T$

$$\boxed{T = U_0 S_0^{-1/2} V_0 \Lambda_0^{1/2}}$$

BALANCED MODEL REDUCTION

MOORE TAC '81
MULLIS ROBERTS IEEE TCAS '76
PERNEBO SILVERMAN TAC '82
AL-SAGGAF, FRANKLIN Proc. CDC 1986

Let $\Sigma = (A, B, C, D)$ Be in BALANCED FORM and Partition

$$A = \begin{array}{c} r \\ n-r \end{array}\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \qquad B = \begin{array}{c} r \\ n-r \end{array}\left[\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right] \qquad C = \begin{array}{c|c} [\,C_1 & C_2\,] \\ r & n-r \end{array}$$

$$P = APA^T + BB^T \qquad P = A^T P A + C^T C \qquad P = \operatorname{diag}\{\sigma_1, \dots, \sigma_n\}$$

$$P = \left[\begin{array}{c|c} P_1 & 0 \\ \hline 0 & P_2 \end{array}\right] \qquad \sigma_r \gg \sigma_{r+1}$$

$\Sigma_r := (A_{11}, B_1, C_1, D)$ = BALANCED TRUNCATION

Theorem (Pernebo Silverman) $[\sigma_r > \sigma_{r+1}]$

A_{11} stable, (A_{11}, B_1) Reach. (A_{11}, C_1) Observ.

Theorem (ENNS, GLOVER)

$$G(z) = C(zI - A)^{-1}B + D$$
$$G_r(z) = C_1(zI - A_{11})^{-1}B_{11} + D$$

$$\| G(z) - G_r(z) \|_\infty \leq 2 \sum_{i=r+1}^{M} \sigma_i \qquad \| G \|_\infty := \sup_\omega \sigma_{\max}(G(j\omega))$$

RELATED CONCEPTS:

- Optimal Hankel Norm Approximation

  - DEF (Hankel norm)

$$G(z) = C(zI-A)^{-1}B + D$$

$$\|G\|_{\mathcal{H}} := \sigma_1 = \sigma_{max}(\mathcal{H}) = \sigma_{max}(\mathcal{O}_\infty \mathcal{R}_\infty)$$

$$= \sigma_{max}^{1/2}(\Pi \Omega)$$

  - DEF (Optimal Hankel norm approximation)

$$\hat{G}_r^{\mathcal{H}}(z) = \underset{\substack{G_r :\\ \text{McMillan deg } G_r \leq r}}{\arg\min} \|G - G_r\|_{\mathcal{H}}$$

  - **Thm** $\|G - G_r\|_{\mathcal{H}} \geq \sigma_{r+1}$ $(\|G\|_\infty < M)$

(Adamyan - Arov - Krein AAK)

HANKEL OPERATOR

$$\underbrace{\begin{bmatrix} u(-1)\\ u(-2)\\ \vdots \\ u(-k) \end{bmatrix}}_{\triangleq\ \underline{u}} \xmapsto{\ \mathcal{H}\ } \underbrace{\begin{bmatrix} y(0)\\ y(1)\\ \vdots \\ y(k) \end{bmatrix}}_{\triangleq\ \underline{y}}$$

$$\boxed{\underline{y} = \mathcal{H}\,\underline{u} = \mathcal{O}_\infty \mathcal{R}_\infty \underline{u}}$$

(Input output map)

Invariant for change of BASIS.

$$\mathcal{H} = U_{\mathcal{H}} \underset{\substack{\uparrow \\ \text{Invariant}}}{S_{\mathcal{H}}} V_{\mathcal{H}}^T = \underset{\mathcal{O}}{\left(U_{\mathcal{H}} S_{\mathcal{H}}^{1/2}\right)}\underset{\mathcal{R}}{\left(S_{\mathcal{H}}^{1/2} V_{\mathcal{H}}^T\right)}$$

$$\mathcal{O}^T\mathcal{O} = S_{\mathcal{H}} \qquad \mathcal{R}\mathcal{R}^* = S_{\mathcal{H}}$$

$$(= \Omega) \qquad\qquad (= \Pi)$$

REMARK: $\sigma_i(\mathcal{H}) = \lambda_i^{1/2}(\mathcal{H}\mathcal{H}^*) = \lambda_i^{1/2}(\mathcal{O}_\infty \mathcal{R}_\infty \mathcal{R}_\infty^* \mathcal{O}_\infty^*)$

$$= \lambda_1^{1/2}(\mathcal{O}_\infty^* \mathcal{O}_\infty \mathcal{R}_\infty \mathcal{R}_\infty^*) = \lambda_i^{1/2}(\Omega\,\Pi)$$

THE HO-KALMAN ALGORITHM

(PARTIAL REALIZATION PROBLEM)

$M(0)\quad M(1)\ \dots\quad M(T) \longleftarrow$ MARKOV PARAMETERS (IMPULSE RESPONSE)

$$\mathcal{H}_{\alpha,\beta} = \begin{bmatrix} M(1) & M(2) & \dots & M(\beta) \\ M(2) & M(3) & & M(\beta+1) \\ \vdots & & & \\ M(\alpha) & M(\alpha+1) & & M(\beta+\alpha-1) \end{bmatrix}$$

$M(k) = CA^{k-1}B$ if $M(k)$ is the impulse responce of $\Sigma = (A,B,C,D)$

$\exists!$ A,B,C s.t.

$$\mathcal{H}_{\alpha,\beta} = \mathcal{O}_\alpha \mathcal{R}_\beta \qquad \mathcal{O}_\alpha = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\alpha-1} \end{bmatrix} \qquad \mathcal{R}_\beta = \begin{bmatrix} B & AB & \dots & A^{\beta-1}B \end{bmatrix}$$

$$\Updownarrow$$

Iff. $\operatorname{rank} \mathcal{H}_{\alpha,\beta} = \operatorname{rank} \mathcal{H}_{\alpha+1,\beta} = \operatorname{rank} \mathcal{H}_{\alpha,\beta+1}$

(rank n)

Algorithm

1) FACTORIZE $\mathcal{H}_{\alpha,\beta}$ (E.Q. VIA SVD. $U_n S_n V_n^T$)

2) Set $\hat{\mathcal{O}}_\alpha = U_n S_n^{1/2}$ $\quad \hat{\mathcal{R}}_\alpha := S_n^{1/2} V_n^T$

3) $C := \left[\hat{\mathcal{O}}_\alpha\right]_{1:m;}$ $\quad B = \left[\hat{\mathcal{R}}_\alpha\right]_{:,1:p}$

$$A = \hat{\mathcal{O}}_\alpha^{-L}\, \overline{\mathcal{H}}_{\alpha,\beta}\, \hat{\mathcal{R}}_\beta^{-R}$$

where

$$\bar{H}_{\alpha,\beta} = \begin{bmatrix} M(2) \cdots & M(\beta+1) \\ \vdots & \\ & \\ M(\alpha+1) & M(\alpha+\beta) \end{bmatrix} \quad \left(= O_\alpha A R_\beta \right)$$

Ho - KALMAN	1966
Tether	1970
ZEIGER McEWEN	19

MODELS FOR RANDOM SIGNALS

$\{y(t)\}_{t\in\mathbb{Z}} \in \mathbb{R}^m$ stochastic process, (Assume always zero mean $\mathbb{E}\, y(t) = 0$)

$(t \in [t_0, +\infty))$

WHICH IS THE STATISTICAL DESCRIPTION OF THE OUTPUT OF A LINEAR SYSTEM?

$$\begin{cases} x(t+1) = A x(t) + B w(t) \\ y(t) = C x(t) + D w(t) \end{cases} \qquad \begin{matrix} x(t_0) = x_0 \\ t \geq t_0 \end{matrix} \qquad \begin{matrix} \mathbb{E}\, x_0 = 0 \\ \mathrm{Var}\, x_0 = \Sigma_0 \end{matrix}$$

$w(t) \in \mathbb{R}^p$ white noise

$$\mathbb{E}\, w(t) w^T(s) = I_p\, \delta(t-s)$$

$$\mathbb{E}\, x_0 w^T(t) = 0 \quad \forall\, t \geq t_0$$

1) OUTPUT STATISTICS?

2) MINIMALITY ?

STATE AND OUTPUT COVARIANCES

$$\begin{cases} x(t+1) = Ax(t) + Bw(t) \\ y(t) = Cx(t) + Dw(t) \end{cases} \qquad \begin{array}{l} t \geq t_0 \\ \mathbb{E}\, x_0 x_0^T = \Sigma_0 \\ x_0 = x(t_0) \end{array}$$

$$\Sigma_x(t,s) := \mathbb{E}\; x(t)\, x^T(s) = \begin{cases} A\; \mathbb{E}\; x(t-1)\, x^T(s) & t > s \\ \Sigma_x(s,s) & t = s \\ \Sigma_x^T(s,t) & t < s \end{cases}$$

$$\Sigma_x(t+1,t+1) = A\,\Sigma_x(t,t)\,A^T + BB^T$$

$$\Rightarrow \quad \Sigma_x(t,t) = A^{t-t_0}\,\Sigma_0\,(A^T)^{t-t_0} + \sum_{i=0}^{t-t_0-1} A^i BB^T (A^T)^i$$

$$\Rightarrow \quad \Sigma_x(t,s) = A^{t-s}\,\Sigma_x(s,s)$$

$$\Sigma_y(t,s) = \begin{cases} \text{[illegible]} \\ CA\Sigma_x(t-1,s)C^T + CA\mathbb{E}[x(t-1)w^T(s)]D^T & t > s+1 \\ C\cancel{A}\cdot[A\Sigma_x(s,s)C^T + BD^T] & t = s+1 \\ C\,\Sigma_x(s,s)\,C^T + DD^T & t = s \end{cases}$$

$$\Sigma_y(t,s) = \begin{cases} CA^{t-s-1}\,G(s) & t \geq s+1 \\ C\,\Sigma_x(s,s)\,C^T + DD^T & t = s \end{cases}$$

OSS: $\mathbb{E}\; x(t-1)\,w^T(s)\,D^T$

$$\hookrightarrow = \begin{cases} A\;\mathbb{E}\,x(t-2)\,w^T(s)\,D^T & t > s+1 \\ BD^T & t = s+1 \end{cases}$$

FOR $t \gg t_0$?

Assume A stable $\left(|\lambda(A)| < 1 \right)$

① $\Sigma_x(t) = \sum_{i=0}^{+\infty} A^{i} BB^{T} (A^{T})^{i}$ DOES NOT DEPEND ON t!

$$\hookrightarrow = \boxed{\Sigma_x = A \Sigma_x A^{T} + BB^{T}}$$ (Lyapunov equation)

$$\Sigma_x(t,s) = A^{t-s} \Sigma_x(s,s) = A^{t-s} \Sigma_x = \Sigma_x(t-s)$$

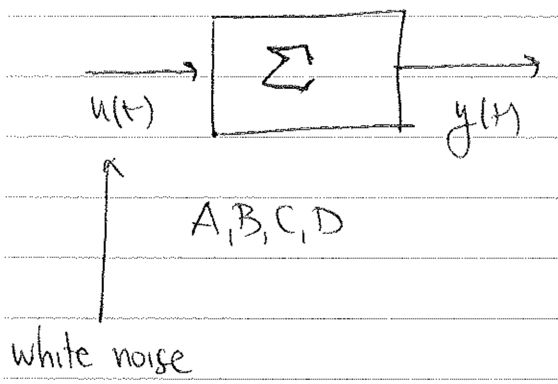
$\uparrow$ Depends only upon difference!

$$\Sigma_Y(t,s) \substack{\neq \\ =} \begin{cases} CA^{t-s-1} G \\ C \Sigma_x C^{T} + DD^{T} \end{cases} \qquad G = A \Sigma_x C^{T} + BD^{T}$$

$= \Sigma_Y(t-s)$ Depends only upon $t-s$

$$|\lambda(A)| < 1 \implies \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \longrightarrow$$ WIDE SENSE (Asymtotically) STATIONARY

COVARIANCE FROM STATE SPACE



$$\Lambda_y(\tau) = \begin{cases} \Lambda(0) & \tau = 0 \\ C A^{\tau-1} G & \tau > 0 \end{cases}$$

$$G = APC^T + BD^T \qquad (= \bar{C}^T)$$

$$P = APA^T + BB^T$$

Th: WIENER - KINTCHINE

$$\Phi(z) = \sum_{\tau=-\infty}^{+\infty} \Lambda(\tau) z^{-\tau} = W(z) W^T(1/z)$$

$$W(z) = C(zI-A)^{-1}B + D$$

Question: is Σ a minima model for $\Phi_y(z)$?

EXAMPLE)

$$\begin{cases} x(t+1) = a x(t) + \left(a - \frac{1}{a} \right) u(t) & |a| < 1 \\ y(t) = x(t) + u(t) \end{cases}$$

$$W(z) = \frac{a - 1/a}{z - a} + 1 = \frac{a - 1/a + z - a}{z - a} = \frac{z - 1/a}{z - a}$$

$$W(z) \cdot W^T(1/z) = \frac{z - 1/a}{z - a} \; \frac{z^{-1} - 1/a}{z^{-1} - a} = \frac{z - 1/a}{z - a} \cdot \frac{1 - z/a}{1 - za} \; \frac{\boxed{a}}{a} \cdot \frac{1/a}{\boxed{1/a}}$$

$$= \frac{\cancel{(z - 1/a)} \; \cancel{(z - a)}}{\cancel{(z - a)} \; \cancel{(z - 1/a)}} \cdot \frac{1}{a^2} = \frac{1}{a^2}$$

$\begin{pmatrix} A & B & C & D \end{pmatrix} = \begin{pmatrix} 2 & 2-1/2 & 1 & 1 \end{pmatrix}$ is Minimal in the deterministic sense but not in the "stochastic" sense

CLASS OF MODELS FOR (WIDE-SENSE) STATIONARY SIGNALS

1) STATE SPACE

$$x(t+1) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t)$$

$$W(z) = C(zI-A)^{-1}B + D$$

2) (VECTOR) ARMA

$$\sum_{k=0}^{n_y} D_k\, y(t-k) = \sum_{k=0}^{n_u} N_k\, u(t-k) \qquad D_0 = I$$

$$\left(W(z) = \left[\sum_{k=0}^{n_y} D_k z^{-k} \right]^{-1} \left[\sum_{k=0}^{n_u} N_k z^{-k} \right] \right.$$

3) SPECTRUM

$$\Phi(z) = \underbrace{W(z)}_{\uparrow} W^T(1/z)$$

Spectral factor

STOCHASTIC REALIZTION PROBLEM

(= Models from Covariance)

$\Sigma_y(k)$, $k = 0, 1, \dots$ <u>GIVEN</u>

FIND: (A, B, C, D) minimal s.t.

$$u(k) \rightarrow \boxed{(A,B,C,D)} \rightarrow y(k) \leftarrow \text{ has covariance } \Sigma_y(k) \ \forall k$$

SOLUTION:

$$\Phi_y(z) = \sum_{k=-\infty}^{+\infty} \Lambda_y(k) z^{-k} = \overbrace{\sum_{k=-\infty}^{-1} \Lambda_y(k) z^{-k} + \frac{1}{2}\Lambda(0)}^{\Phi_+^T(1/z)} + \underbrace{\frac{1}{2}\Lambda(0) + \sum_{k=1}^{+\infty} \Lambda(k) z^{-k}}_{\Phi_+(z)}$$

$\Phi_+(z)$ is the transfer function of a "filter" with impulse response $\left(\frac{1}{2}\Lambda(0), \Lambda(1), \Lambda(2), \dots \Lambda(k), \dots \right)$

OSS: $\Phi_y(e^{j\theta}) \geq 0 \iff$ $\Phi_+(z)$ is POSITIVE REAL

DEFINE:

$$\mathcal{H}_y := \begin{bmatrix} \Lambda_y(1) & \Lambda_y(2) & \cdots \\ \Lambda_y(2) & \Lambda_y(3) & \cdots \\ \vdots & & \\ & & \end{bmatrix} \qquad \Sigma_y(k) = CA^{k-1}\bar{C}^T$$

$$\bar{C}^T = APC^T + BD^T$$

$$\Rightarrow \quad \mathcal{H}_y = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ \vdots \end{bmatrix}}_{\mathcal{O}} \underbrace{\begin{bmatrix} \bar{C}^T & A\bar{C}^T & \cdots \end{bmatrix}}_{\mathcal{C}}$$

⇒ GIVEN $\{\Lambda(0)\ \Lambda(1) \ldots \ \Lambda(k) \ldots\}$ FIND !

$(A, C, \bar{C}, \Lambda(0))$ s.t. $\Lambda(i) = CA^{i-1}\bar{C}^T$

PROBLEM: $\Phi_+(z) = \frac{1}{2}\Lambda(0) + C(zI-A)^{-1}\bar{C}$

has to be positive real.

(OK if $\Lambda(k)$ given $\forall$ k ......)

GIVEN $\left(A, C, \bar{C}, \Lambda(0)\right)$, find B, D

OBSERVATION.

$$\phi(z) = W(z)W^T(1/z)$$

$$= \begin{bmatrix} C(zI-A)^{-1} & I \end{bmatrix} \underbrace{\begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B^T & D^T \end{bmatrix}}_{\geq 0} \begin{bmatrix} (z^{-1}I-A^T)^{-1}C^T \\ I \end{bmatrix}$$

$$BB^T = P - APA^T$$

$$BD^T = \bar{C}^T - APC^T$$

$$DD^T = \Lambda(0) - CPC^T$$

$$\Rightarrow \begin{bmatrix} P - APA^T & \bar{C}^T - APC^T \\ \bar{C} - CPA^T & \Lambda(0) - CPC^T \end{bmatrix} \geq 0 \qquad (*)$$

POSITIVE REAL LEMMA (KYP) [KALMAN - YAKUBOVICH - POPOV]

$\left(A, C, \bar{C}, \frac{\Lambda(0)}{2}\right)$ is Positive Real (i.e. $\phi_+(z) = C(zI-A)^{-1}\bar{C} + 1/2\,\Lambda(0)$ is positive real.)

iff. $(*)$ Admits solutions $P = P^T \geq 0$

INTERNAL REALIZATIONS $(\dim u(t) = \dim y(t))$

$$\Leftrightarrow \quad \operatorname{rank} \begin{bmatrix} P - APA^T & \bar{C}^T - APC^T \\ \bar{C} - CPA^T & \Lambda(0) - CPC^T \end{bmatrix} = m$$

(since $\Lambda(0) - CPC^T$ invertible)

$$\underset{(A.R.E)}{\Leftrightarrow} \quad \boxed{P - APA^T - (\bar{C}^T - APC^T)(\Lambda(0) - CPC^T)^{-1}(\bar{C} - CPA^T)^T = 0}$$

$(**)$

Algebraic Riccati Equation

- Solutions to $(**)$

$\forall \; P = P^T \geq 0$ sol. of $(**)$, $\quad P_- \leq P \leq P_+$

P_- and P_+ are "special" solutions of $(**)$

$\uparrow \qquad\qquad \uparrow$

"minimal" "maximal"

SPECIAL SOLUTIONS

P_- (minimal) $\Rightarrow \begin{bmatrix} B_- \\ D_- \end{bmatrix} \Rightarrow W_-(z) = C(zI-A)^{-1}B_- + D_-$

$W(z)$ is Minimum Phase ($\equiv$ all zeros inside the unit disc)

P_+ (maximal) $\Rightarrow \begin{bmatrix} B_+ \\ D_+ \end{bmatrix} \Rightarrow W_+(z) = C(zI-A)^{-1}B_+ + D_+$

$W_+(z)$ is "maximum Phase" ($\equiv$ all zeros outside the unit disc)

MEANING:

$y(t) = W_-(z)\, u(t)$ (stable)

$u(t) = W_-^{-1}(z)\, y(t)$ (stable)

THE KALMAN FILTER

$$x(t+1) = Ax(t) + \omega(t) \qquad \mathrm{Var.}\begin{Bmatrix} \omega(t) \\ v(t) \end{Bmatrix} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}$$

$$y(t) = Cx(t) + v(t)$$

$$\hat{x}(t+1) = A\hat{x}(t) + K(t)e(t) \qquad t \geq 0$$

$$y(t) = C\hat{x}(t) + e(t) \qquad \hat{x}(0) = 0$$

$$\tilde{P}(0) = P$$

$$K(t) = [A\tilde{P}(t)C^T + S][C\tilde{P}(t)C^T + R]^{-1}$$

$$\tilde{P}(t+1) = A\tilde{P}(t)A^T - K(t)\Lambda(t)K^T(t) + Q$$

$$P = APA^T + Q \qquad \hat{P}(t) := P - \tilde{P}(t)$$

$$(\hat{P}(t) = \mathrm{Var}\{\hat{x}(t)\})$$

$$\hat{P}(t+1) = A\hat{P}(t)A^T + K(t)\Lambda(t)K^T(t)$$

$$\Downarrow$$

$$\hat{P}(t+1) - A\hat{P}(t)A^T - K(t)\Lambda(t)K^T(t) = 0$$

$$\lim_{t \to +\infty} \hat{P}(t) = \hat{P}_\infty$$

$$\hat{P}_\infty - A\hat{P}_\infty A^T - K_\infty \Lambda_\infty K_\infty^T = 0$$

$$K_\infty = [A\tilde{P}_\infty C^T + S][C\tilde{P}_\infty C^T + R]^{-1}$$

$$= [A(P-\hat{P}_\infty)C^T + S][C(P-\hat{P}_\infty)C^T + R]^{-1} = (\bar{C}^T - A\hat{P}_\infty C^T)(\Lambda(0) - C\hat{P}_\infty C^T)^{-1}$$

$$\begin{cases}\hat{x}(t+1) = A\hat{x}(t) + K_\infty e(t)\\ y(t) = C\hat{x}(t) + e(t)\end{cases} \iff \begin{cases}\hat{x}(t+1) = (A - K_\infty C)\hat{x}(t) + K_\infty y(t)\\ e(t) = -C\hat{x}(t) + y(t)\end{cases}$$

$$y(t) = \underset{\substack{\uparrow \\ \text{analytic } (|\lambda(A)|<1)}}{W(z)}\, e(t) \qquad\qquad e(t) = \underset{\substack{\uparrow \\ \text{analytic} \\ (|\lambda(A-K_\infty C)|<1)}}{W^{-1}(z)}\, y(t)$$

- The steady state Kalman filter computes the minimum phase spectral factor $W_-(z)$ of $\Phi_y(z)$
- The (steady state) Kalman gain K_∞ is given by $K_\infty = (\bar{C}^T - A\hat{P}_\infty C^T)(\Lambda(0) - C\hat{P}_\infty C^T)^{-1}$ where $\hat{P}_\infty$ is the minimal solution of ARE

$$\hat{P}_\infty - A\hat{P}_\infty A^T - (\bar{C}^T - A\hat{P}_\infty C^T)(\Lambda(0) - C\hat{P}_\infty C^T)^{-1}(\bar{C}^T - A\hat{P}_\infty C^T)^T = 0$$

- $\hat{P}_\infty$ (minimal solution of ARE) is stabilizing i.e. $|\lambda(A - K_\infty C)| < 1$

BACKWARD MODELS AND "MAXIMUM PHASE"

Backward Kalman filter $\longleftrightarrow$ $W_+(z)$

(maximum phase)

- MAIN "CONCEPTUAL" STEPS IN SUBSPACE IDENTIFICATION

1. Partial covariance extension

$$\begin{tikzcd}
(\Lambda(0), \dots, \Lambda(k)) \quad k \text{ finite} \arrow[d] \\
(A, C, \bar{C}, \tfrac{1}{2}\Lambda(0)) \arrow[d, "KYP"] \\
B, D \arrow[d] \\
\underset{\sim}{W}(z) = \text{spectral factor.}
\end{tikzcd}$$

2. Replace covariances with <u>sample covariances</u>.

$$(\hat{\Lambda}(0), \dots, \hat{\Lambda}(k))$$

3. From DATA $\{y(0), \dots, y(T)\}$

compute a basis for $X(t)$ $\longrightarrow$ compute

$$(A, C, \bar{C}, \tfrac{1}{2}\Lambda(0)) \xrightarrow[KYP]{} (B, D) \longrightarrow W(z)$$

~~[illegible] the~~

$[\alpha_1, \dots, \alpha_N] \longrightarrow$ To be determined

$\uparrow$

As many parameters as data points. (N)

EXAMPLE 3

$$y(t_i) = f(t_i) \overset{\perp\!\!\!\perp}{+} n(t_i) \qquad t_i \in \{t_1, \dots, t_N\}$$

$$n(t_i) \perp n(t_j) \quad i \neq j$$

$$\mathbb{E}\, n^2(t_i) = \sigma^2$$

GAUSSIAN PROCESS PRIOR ON $f(t)$

$$f(t) \sim \mathcal{N}(0, K(\cdot,\cdot))$$

$$K(t,s) = \mathbb{E}[f(t) f^T(s)]$$

$$\hat{f}(t) := \mathbb{E}[f(t) \mid y(t_1), \dots, y(t_N)] \qquad \underline{Y} := \begin{bmatrix} y(t_1) \\ \vdots \\ y(t_N) \end{bmatrix}$$

$$= [\text{cov.}\{f(t), y(t_1)\}, \dots, \text{cov.}\{f(t) y(t_N)\}] \cdot \text{Var.}\{\underline{Y}\}^{-1} \cdot Y$$

$$\text{cov.}\{f(t), y(t_i)\} = K(t, t_i)$$

Kronecker Delta.

$$(\text{Var.}\, \underline{Y})_{i,j} = K(t_i, t_j) + \sigma^2 \delta_{ij}$$

DEFINE! $\quad \hat{\underline{\alpha}} := \text{Var.}\{\underline{Y}\}^{-1} Y = \begin{bmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_N \end{bmatrix}$

$$\Rightarrow \boxed{\hat{f}(t) = [K(t,t_1) \dots K(t,t_N)] \begin{bmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_N \end{bmatrix} = \sum_{i=1}^{N} \hat{\alpha}_i K(t,t_i)}$$

The estimator

$$\hat{f}(t) = \sum_{i=1}^{N} \hat{\alpha}_i \, K(t, t_i) \quad ;$$

(Tikhonov)
Regularization
Framework

is also known as <u>Regularization Network</u>

Rich Theory behind : GAUSSIAN PROCESSES / Reproducing Kernel Hilbert Spaces (RKHS)

BAYES FRAMEWORK

- We shall come back to this at the end of the course.

PARAMETRIC STATISTICS (Point Estimators)

$\underline{Y} := [Y_1, \ldots, Y_N]$ random Variables $\qquad M(\theta) := P_\theta(\underline{y}) \quad \theta \in \Theta$

↓ Experiment $\qquad\qquad$ ↑ Parametric Model Class

(ES: JOINT PROBABILITY)

$\underline{y} := [y_1, \ldots, y_N]$

$T(\underline{Y}) := T(Y_1, \ldots, Y_N)$ STATISTIC ((Measurable) Function of Y which DOES NOT depend upon unknown parameters)

$\hat{\theta}_N(\underline{Y})$ ESTIMATOR : A statistic use to estimate the parameter θ

↓

It is a random Variable.

UNBIASED: $\mathbb{E}[\hat{\theta}_N(\underline{Y})] = \theta \quad \forall \theta$

CONSISTENT: $\lim_{N \to \infty} \hat{\theta}_N(\underline{Y}) \text{ ``}=\text{'' } \theta$

↑ (In probability, a.s, in the q.m etc...)

$\text{Var}\{\hat{\theta}_N(Y)\} = ?$ ← IDEALLY WOULD LIKE TO BE SMALL!

↓

HOW SMALL CAN THIS BE?

CRAMÈR - RAO LOWER BOUND

(Under suitable Regularity conditions ...)

Let $g(\theta) := \mathbb{E}[\hat{\theta}]$ and $I(\theta) := \mathbb{E}\left[\frac{\partial \log p_\theta}{\partial \theta} \cdot \frac{\partial \log^T p_\theta}{\partial \theta}\right]$

$$\text{Var}\{\hat{\theta}\} \geq \dot{g}(\theta)\, I^{-1}(\theta)\, \dot{g}^T(\theta)$$

Where $[\dot{g}(\theta)]_{i,j} := \frac{\partial g_i}{\partial \theta_j}$

Pf: $s_\theta := \frac{\partial \log p_\theta}{\partial \theta}$ (SCORE)

$$\mathbb{E}\, s_\theta = \int s_\theta\, p_\theta(\underline{y})\, d\underline{y} = \int \frac{\partial p_\theta(y)}{\partial \theta}\, dy$$

Need Regularity!

$$\overset{!}{=} \frac{\partial}{\partial \theta} \underbrace{\int p_\theta(y)\, dy}_{=1} = 0$$

Compute $\hat{E}[\hat{\theta} \mid s_\theta] := \text{cov}\{\hat{\theta}, s_\theta\}\, \text{Var}\{s_\theta\}^{-1} s_\theta$

Then: $\underbrace{\text{Var}\{\hat{\theta}\} - \text{cov}\{\hat{\theta}, s_\theta\}\, \text{Var}\{s_\theta\}^{-1}\, \text{cov}^T\{\hat{\theta}, s_\theta\}}_{\text{This is a Variance and Hence it is positive semidefinite!}} \geq 0$

- Since s_θ has zero mean

$$\hookrightarrow \text{cov}\{\hat{\theta}, s_\theta\} = \mathbb{E}[\hat{\theta}\, s_\theta^T] = \int \hat{\theta} \cdot \frac{\partial \log^T p_\theta}{\partial \theta} \cdot p_\theta(y)\, dy$$

$$= \int \hat{\theta}(y)\, \frac{\partial p_\theta^T(y)}{\partial \theta}\, dy = \frac{\partial^T}{\partial \theta}\left[\int \hat{\theta}(y)\, p_\theta(y)\, dy\right]$$

$$= \left[\frac{\partial g(\theta)}{\partial \theta}\right]$$

$$\text{Var.}\{s_\theta\} = \mathbb{E}\left[\frac{\partial \log p_\theta}{\partial \theta}\left(\frac{\partial \log p_\theta}{\partial \theta}\right)^T\right] = I(\theta)$$

$$\Rightarrow \text{Var.}\{\hat{\theta}\} - \mathring{g}(\theta)\, I^{-1}(\theta)\, \mathring{g}^T(\theta) \geq 0 \qquad \boxed{\text{C.V.D.}}$$

DEF: if $g(\theta) = \theta$ and

$$\text{Var.}\{\hat{\theta}\} = \mathring{g}(\theta)\, I^{-1}(\theta)\, \mathring{g}^T(\theta)$$

$$\left(\text{i.e.} \quad \text{Var.}\{\hat{\theta}\} = I^{-1}(\theta)\right)$$

$\hat{\theta}$ IS SAID TO BE EFFICIENT

HOW DOES ONE CHOOSE AN ESTIMATOR ?

- Optimality. (Minimum Variance Unbiased Estimator)

  ↳ Based on the Theory of Sufficient Statistics .....

  ... Often hard to find ...

- MAXIMUM LIKELYHOOD

$$\hat{\theta}^{ML} \in \arg\sup_{\theta} P_\theta(y_1, \dots, y_N)$$

CONS:
1) Not always unbiased (often!)
2) Difficult (non-convex, high dimens.) optimization problem. (LOCAL MINIMA/MAXIMA !)

PROS: (Under mild conditions)

1) Consistent. (a.s.)
2) Asymptotically Normal
3) Asymtotically Efficient.

2)+3)

$$\sqrt{N}\left(\hat{\theta}_N^{ML} - \theta\right) \underset{N \to \infty}{\overset{\substack{\text{conv. in Legge} \\ \uparrow}}{\rightsquigarrow}} \mathcal{N}\Big(0, \underbrace{I^{-1}(\theta)}_{\text{Asymptotic Variance}}\Big)$$

(≡ Variance of the Asymtotic Distribution)

$$\boxed{\text{Asympotic Efficiency}} \longrightarrow |||$$

CRAMÈR - RAO LOWER BOUND (Normalized x unit observation)

- In general: optimization BASED approaches

  1. Choose $J(\theta)$ "cost function" which says How well the parameter θ describes the data $(y_1, \dots, y_N)$

  2. $\hat{\theta}^J \in \arg\sup_{\theta} J(\theta)$

     → Often same problems as ML (non convex etc...)

     → Properties depend upon the choice of $J(\theta)$

- METHOD OF MOMENTS (No optimality property in general.)

  1) Compute Sample moments

  2) Choose $\hat{\theta}$: Moments $M(\hat{\theta}) =$ Sample Momements

     $\overset{\shortparallel}{\text{(population moments)}}$

(Hogg - Craigg)

SUFFICIENT STATISTIC

DEF: $T(\underline{Y})$ is a sufficient statistic for θ

if $p(\underline{y} \mid T(\underline{y}))$ DOES NOT DEPEND UPON θ

Intuitively: once $T(y)$ is known, there is no information left in $\underline{Y}$ to determine θ

FACTORIZATION CRITERION (NEYMAN - PEARSON)

$$T(\underline{Y}) \text{ sufficient} \iff p_\theta(y_1,\dots,y_N) = h(\underline{y}) \cdot f_\theta(T(\underline{y}))$$

EXAMPLE $Y \sim B(\theta)$ $Y = \begin{cases} 0 & P[Y=0] = 1-\theta \\ 1 & P[Y=1] = \theta \end{cases}$

$Y_1, \dots, Y_N$ i.i.d $\sim B(\theta)$

$$p_\theta(y_1, \dots y_N) = \prod_{i=1}^{N} \theta^{y_i} (1-\theta)^{1-y_i} = \theta^{\sum y_i} (1-\theta)^{\sum(1-y_i)}$$

$$= \underbrace{(1-\theta)^N}_{h(\underline{y})} \cdot \underbrace{\left(\frac{\theta}{1-\theta}\right)^{\sum_i y_i}}_{= f_\theta(T(\underline{y}))}$$

$\sum_{i=1}^{N} Y_i = T(\underline{Y})$ is a sufficient statistics

With some work (Minimality and Completeness) (*)

can show that if $T(\underline{Y})$ is a minimal, complete sufficient statistic, then $T(\underline{Y})$ is an optimal (MVUE) estimator for $g(\theta) := \mathbb{E}\, T(\underline{Y})$ $\left[\begin{array}{l}\text{Rao-Blackwell} \\ \text{Lehman - sheffè}\end{array}\right]$

* * *

$T(\underline{Y}) := \sum_{i=1}^{N} Y_i$ is a minimal complete statistic for θ in the Bernoulli model

$\Downarrow$

$T(\underline{Y})$ is an optimal estimator for $\mathbb{E}\, T(\underline{Y}) = N\theta$

$\Rightarrow \frac{1}{N} T(\underline{Y})$ is an optimal estimator for $\mathbb{E}\, \frac{1}{N} T(\underline{Y}) = \theta$

* * *

(*) MINIMALITY

$T(Y)$ minimal sufficient iff:

1) $T(Y)$ sufficient
2) $\forall\ T_1(Y)$ sufficient, $\exists\, f(\cdot) : T(Y) = f(T_1(Y))$

(*) COMPLETENESS

$T(Y)$ is complete if $\forall\ f(\cdot)$ measurable s.t.

$$\mathbb{E}\left[f(T(Y))\right] = 0 \implies \mathbb{P}_\theta\left(f(T(Y)) = 0\right) = 1 \quad \forall\, \theta$$

Other Useful facts:

DEF: ANCILLARY STATISTIC :

Statistic whose distribution DOES NOT depend on the parameters (unknown)

EX: $Y_1, \dots, Y_N$ i.i.d. $Y_i \sim N(\overset{\text{Unknown}}{\theta}, \overset{\text{Known}}{\sigma^2})$ (1)

$$S^2 := \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \qquad \bar{Y} := \frac{1}{N} \cdot \sum_{i=1}^{N} Y_i$$

S^2 is Ancillary $\left(\frac{(N-1)S^2}{\sigma^2} \sim \chi^2(N-1) \right)$

BASU'S THEOREM: [SANKHYA , SER A, 15 (1955)]

Any complete sufficient statistic is independent of any ancillary statistic

EX: (CONT'D) $\bar{Y}$ is complete sufficient for θ in (1)
S^2 is Ancillary

$$\Downarrow$$

$\bar{Y}, S^2$ are independent.

EXAMPLE 2

$Y \sim \mathcal{U}[0,\theta]$ $\quad Y_1,\dots,Y_N$ i.i.d. $\sim \mathcal{U}[0,\theta]$ $\quad p_\theta(y) = \frac{1}{\theta}\cdot \chi^{y}_{[0,\theta]}$

$$\chi^{y}_{[0,\theta]} = \begin{cases} 1 & y \in [0,\theta] \\ 0 & y \notin [0,\theta] \end{cases}$$

$$p_\theta(y_1,\dots,y_N) = \frac{1}{\theta^N}\prod_{i=1}^{N} \chi^{(y_i)}_{[0,\theta]} = \frac{1}{N}\cdot \chi^{(\max y_i)}_{[0,\theta]}$$

$T(\underline{Y}) = \max\{Y_i\}$ is a sufficient statistic

$\hat{\theta}_{ML} = \max\{Y_i\}$ (PROVE!)

(However $\mathbb{E}\,\hat{\theta}_{ML} \neq \theta$)

HOMEWORK: show that $\hat{\theta}_{OPT} = \frac{N+1}{N}\,\hat{\theta}_{ML}$.

$\uparrow$

Optimal Estimator

$\left(= \frac{N+1}{N}\max\{Y_i\}\right)$

EXAMPLE 3 (method of moments)

$y_t - a y_{t-1} = e_t + b e_{t-1}$ $\quad e_t$ white noise $\sim \mathcal{N}(0, \sigma^2)$

$$y_t = \frac{z + b}{z - a} e_t = \left[\frac{(b+a)}{z-a} + 1 \right] e_t$$

STATE SPACE MODEL

NORMALIZED

$\downarrow$

$$x_{t+1} = a x_t + (b+a)\sigma \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0,1)$$
$$y_t = x_t + \sigma \varepsilon_t$$

Compute moments

$\mathbb{E} x_t^2 = p$

STATIONARITY $\quad p = a^2 p + (b+a)^2 \sigma^2 \Rightarrow p = \dfrac{(b+a)^2 \sigma^2}{1-a^2}$

$$\mathbb{E}\, y_t y_t = p + \sigma^2 = r(0)$$

$$\mathbb{E}\, y_{t+1} y_t = a\, \mathbb{E}\, x_t y_t + (b+a)\sigma\, \mathbb{E}\, \varepsilon_t y_t$$

$$= a p + (b+a)\sigma^2 = r(1)$$

$$\mathbb{E}\, y_{t+2} y_t = a\, \mathbb{E}\, x_{t+1} y_t = a \left\{ a\, \mathbb{E}\, x_t y_t + (b+a)\sigma\, \mathbb{E}\, \varepsilon_t \varepsilon_t \right\}$$

$$= a\left(a p + (b+a)\sigma^2 \right) = a\, r(1)$$

$$\hat{r}(0) = p + \sigma^2$$
$$\hat{r}(1) = ap + (b+a)\sigma^2$$
$$\hat{r}(2) = a(ap + (b+a)\sigma^2)$$

$$p = \frac{(b+a)^2\sigma^2}{1-a^2}$$

$$\Downarrow$$

$$\left\{ \hat{a} = \frac{\hat{r}(2)}{\hat{r}(1)} \qquad \hat{r}(0) = p + \sigma^2 \qquad \hat{r}(1) = \hat{a}p + (b+\hat{a})\sigma^2 \right.$$

$$\hat{r}(0) = \frac{\left[(b+\hat{a})^2 + \hat{a}^2 + 1\right]\sigma^2}{1-\hat{a}^2} = \left(\frac{b^2 + 2\hat{a}b + 1}{1-\hat{a}^2}\right)\sigma^2$$

$$\hat{r}(1) = \frac{1}{1-\hat{a}^2}\left[\hat{a}(b+\hat{a})^2 + (1-\hat{a}^2)(b+\hat{a})\right]\sigma^2$$

$$= \frac{\left[\hat{a}(b^2 + \cancel{\hat{a}^2} + 2b\hat{a}) + b + \hat{a} - \hat{a}^2 b \cancel{- \hat{a}^3}\right]\sigma^2}{1-\hat{a}^2}$$

$$= \frac{b\hat{a}^2 + \hat{a}b^2 + b + \hat{a}}{(1-\hat{a}^2)}\,\sigma^2$$

May not have a solution!

$\operatorname{var}\{\varepsilon_t\} = 1$

EG: if $y_t = \varepsilon_t + \varepsilon_{t-2}$

$r(0) = 2$
$r(1) = 0$
$r(2) = 1$
$r(k) = 0 \quad k > 2$

$r(2)$ is never $= a \cdot r(1)$ $\forall a$!

PARAMETRIC ESTIMATION

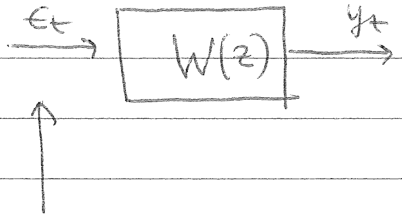
1. DATA $\{y_1, \dots, y_N\}$
2. MODEL CLASS $M(\theta)$ $\theta \in \Theta$
3. CRITERION OF FIT $J(\theta)$

Example: ARMA models

$\{y_t\}_{t \in \mathbb{Z}}$ stationary (ergodic) stochastic process

$\Lambda(\tau) := \mathbb{E}[\, y(t+\tau)\, y^T(t)\,]$ covariance function

$\Phi_y(z) = \sum_{\tau=-\infty}^{+\infty} \Lambda(\tau)\, z^{-\tau}$ $\alpha < |z| < \beta$ $\alpha < 1 < \beta$



WIENER - KINTCHINE

$$\boxed{\Phi_y(z) = W(z)\, W^T(1/z)}$$

ε_t white noise

$\mathbb{E}[\varepsilon_t \varepsilon_s^T] = I\, \delta(t-s)$

$\Phi_\varepsilon(z) = I$

$W(z)$ Rational $\longleftrightarrow$ $\Phi_y(z)$ Rational

$\updownarrow$

Polynomial / State Space

model of finite order

Assume $b = 0$

$$r(0) = \frac{\sigma^2}{1-a^2}$$

$$r(1) = a\, r(0)$$

$$\hat{a} = \frac{\hat{r}(1)}{\hat{r}(0)}$$

$$\hat{\sigma}^2 = \hat{r}(0) - \frac{\hat{r}^2(1)}{\hat{r}(0)}$$

When is there a solution? $\hat{r}(0) \neq 0$ (ovvio).

$|\hat{a}| < 1$ $\quad(\hat{\sigma}^2 > 0)$

$$\hat{r}(0) - \frac{\hat{r}^2(1)}{\hat{r}(0)} > 0 \iff \underbrace{\begin{bmatrix} \hat{r}(0) & \hat{r}(1) \\ \hat{r}(1) & \hat{r}(0) \end{bmatrix} > 0}_{}$$

Estimated covariance is a "valid" covariance (positive definite)

$$\begin{bmatrix} r(0) & r^T(1) \\ r(1) & r(0) \end{bmatrix} = \mathbb{E}\begin{bmatrix} y(t) \\ y(t+1) \end{bmatrix}\begin{bmatrix} y^T(t) & y^T(t+1) \end{bmatrix} > 0$$

2) Choose a parametric class $W_\theta(z)$

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min} \; \underset{\uparrow}{J(\theta)}$$

CRITERION OF FIT

PEM (Prediction Error Methods)

$W_\theta(z) \longleftrightarrow \hat{y}_\theta(t|t-1)$ Best (Linear) Predictor

$(\Phi_y(z,\theta))$

$\varepsilon_\theta(t) := y(t) - \hat{y}_\theta(t|t-1)$ Prediction Error

$$\hat{\theta}^{PEM} = \underset{\theta}{\arg\min} \frac{1}{N} \sum_{t=1}^{N} h\Big[\underbrace{\varepsilon_\theta(t)\, \varepsilon_\theta^T(t)}_{\downarrow \; [m \times 1]\,[1 \times m]} \Big]$$

$h(\cdot) : \mathbb{R}^{m \times m} \longmapsto \mathbb{R}_+$

$h(\cdot)$ is monotone (i.e. $h(Q_1) \geq h(Q_2)$)

iff $Q_1 \geq Q_2$

Typical choices

$h(Q) = Tr\{W \cdot Q\}$

$h(Q) = \det(Q)$

PROBLEMS WITH PEM (As any optimization Based method.)

$J(\theta)$ is a non-convex (concave) function of θ.

a) ⟶ LOCAL MINIMA. → any gradient based method is likely to get stuck in local minima/maxima.

b) Need to find a parametrization θ for the model

↓

THIS IS NOT AN EASY TASK FOR MIMO SYSTEMS (Canonical forms, kronecker indeces...)

IN THIS COURSE WE SHALL SEE A CLASS OF NEW (LAST 2/3 DECADES) METHODS WHICH ARE NOT BASED ON ITERATIVE SEARCH BUT ~~RATHER~~ RELY ONLY ON NUMERICALLY ROBUST TOOLS IN LINEAR ALGEBRA (QR, SVD ....)