Fit

Total Sum of Squares:

$$SS_T = \sum (y - \bar{y})^2$$

Regression Sum of Squares:

$$SS_R = \sum (\hat{y} - \bar{y})^2$$

Holds ↓

$\langle SS_T = SS_E + SS_R \rangle$

*Error of Sum of Squares:

$$SS_E = \sum (y - \hat{y})^2$$

- We want $SS_E \to 0$

Fit

> $R = \sqrt{\dfrac{SS_R}{SS_T}}$ → Perfect Fit: $\underline{\underline{R = 1}}$ $(\hat{y} = y)$

→ Bad Fit: $\underline{\underline{R = 0}}$ $(\hat{y} = \bar{y})$

> **Confidence Interval**

<A>1: $E[\varepsilon] = 0$

<A>2: $\varepsilon(k) \sim \mathcal{N}(0, \sqrt{\Sigma})$

$$\hat{\theta} \sim \mathcal{N}(\theta, \Sigma(\phi^T \phi)^{-1})$$

$$\hat{y} \sim \mathcal{N}(\phi \theta, [I - \phi(\phi^T \phi)^{-1} \phi^T]$$

5% Confidence Interval:

> $\theta_j = \hat{\theta}_j \pm 2\sqrt{\Sigma}\sqrt{d_{jj}}$ (Normal Dist) <A>

23 NOV 15 – y(k)

$$y(k) = \sum_{i=1}^{N} x_i(k)$$

$\xleftrightarrow{N \text{ params}}$

$\updownarrow$ T measurements

> $\underset{\langle T\times 1\rangle}{y} = \underset{\langle T\times N\rangle}{\phi}\,\underset{\langle N\times 1\rangle}{\theta} + \varepsilon$ $\theta = (\theta_1, \ldots, \theta_N)$

> $\hat{\theta} = (\phi^T\phi)^{-1}\phi^T y$ ⟨OLS [illegible] : (Optimal)⟩ OLS minimizes

$$\sum (\underbrace{y(t) - \hat{y}(t)}_{e})^2 \quad \text{if } E[\varepsilon] = 0$$

> $\hat{y} = \phi\,\theta$

$\Sigma = Cov(\varepsilon)$ → "Don't know"

> $\hat{\Sigma} = \dfrac{e^T e}{T-N}$ → Est. of Σ

> $Cov(\hat{\theta}) = \Sigma(\phi^T\phi)^{-1}$

$\approx \hat{\Sigma}(\phi^T\phi)^{-1} = \hat{\Sigma} D$

> Correlation coeff.s : $r_{ij} = \dfrac{d_{ij}}{\sqrt{d_{ii}}\sqrt{d_{jj}}}$

▷ ID of D-Time Model

$$u(t) \rightarrow \boxed{F(s)} \rightarrow y(t)$$

∘ D Model

Linear ←
k: sample index

$$y[k+1] - a_1 y(k) = b_1 u(k) + b_2 u(k-1)$$

$$y[k] + a_1 y[k-1] = b_1 u(k-1) + b_2 u(k-2)$$

Z

$$y(z) + a_1 z^{-1} y(z) = b_1 z^{-1} u(z) + b_2 z^{-1} u(z)$$

$$\frac{y(z)}{u(z)} = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1}}$$

∘ Criteria

Look for the model $(\hat{a}_1, \hat{b}_1, \hat{b}_2)$

$\$$ Best Predict the "One Step Ahead" ⊡

real measurements: $y(k)$, $y(k-1)$, $u(k-1)$, $u(k-2)$

model: $(\hat{a}_1, \hat{b}_1, \hat{b}_2)$

predict ⊡: $\hat{y}(k) = -\hat{a}_1 y(k-1) + \hat{b}_1 u(k-1) + \hat{b}_2 u(k-2)$

Q. Find $(\hat{a}_1, \hat{b}_1, \hat{b}_2)$ $\$$ minimizes $y(k) - \hat{y}(k)$

$$y(k) = \begin{bmatrix} -b(k-1), & u(k-1), & u(k-2) \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix}$$

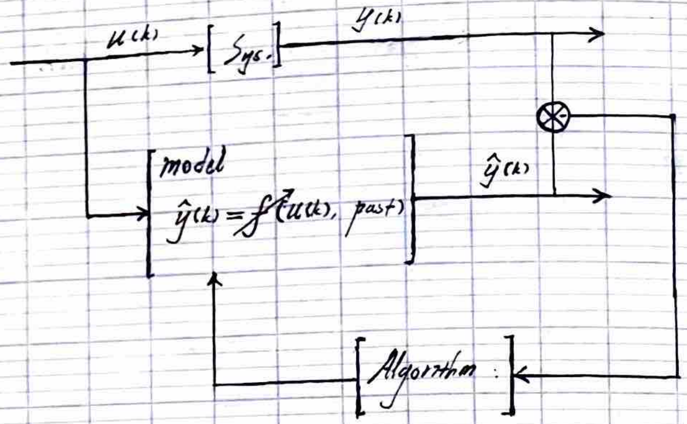
$$= \phi(k)\,\theta(k)$$

↳ θ Lin.

↳ DLS

1. Intro.

> <G> ID [paragraph]



<O> Find Optimal f

<f>

2. Gradient Search

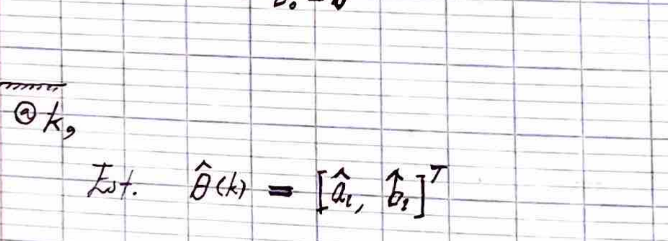
Model: $y(k+1) = -a_1 y(k) + b_1 u(k)$ $\xrightarrow{Z}$ $z\,y(z) = -a_1 y(z) + b_1 u(z)$

$\theta = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$ $\quad y(k+1) = \theta^T \phi(k)$

$\phi = [-y(k), u(k)]$

$$\frac{y(z)}{u(z)} = \frac{b_1 z^{-1}}{1 + a_1 z^{-1}} \qquad \frac{b_1}{z + a_1}$$

($z+1$: past (?))



$\Rightarrow b_0 = 0$

@ k,

Est. $\hat{\theta}(k) = [\hat{a}_1, \hat{b}_1]^T$

we have $y(k)$, $u(k)$

@ $k+1$,
we receive $y(k+1)$

We can predict $\hat{y}(k+1) = -\hat{a}_1 y(k) + \hat{b}_1 u(k)$

$= \hat{\theta}^T(k)\,\phi(k)$

• Compute A-Priori error

$\varepsilon^{\circ}(k) = y(k+1) - \hat{y}(k+1)$

* Choice: <u>Local Criteria</u>

$$\mathcal{I}(k) = \left(y(k+1) - \hat{\theta}^T(k) - \phi(k)\right)^2$$

* We look for

$$\hat{\theta}(k+1) = \hat{\theta}(k) + D$$

$\ni$ $\mathcal{I}$ is minimized

* gradient search

$$D = -F \nabla_{\hat{\theta}} \mathcal{I} = -F \frac{\partial \mathcal{I}}{\partial \hat{\theta}}$$

$\hookrightarrow$ Learning Rate Matrix (>0)

$$= +F \cdot 2\phi(k) \left[y(k+1) - \hat{\theta}^T(k) \phi(k) \right]$$

* Adaptive of $\hat{\theta}$

$$\hat{\theta}(k+1) = \hat{\theta}(k) + F \cdot 2\phi(k) \, \varepsilon_o(k) \qquad \text{with } \varepsilon_o(k) = y(k) - \hat{\theta}^T(k)\phi(k) \qquad \cdots (\mathrm{I})$$

* Compute a-posteriori Error

$$\varepsilon(k) = \hat{y}(k+1) - \hat{\theta}^T(k+1)\phi(k) \quad \longleftrightarrow \quad \varepsilon(k) = \frac{\varepsilon_o(k)}{1 + \phi(k) F \phi^T(k)} \qquad \cdots (\mathrm{II})$$

RECURSIVE
∇ Search.

$$\hat{\theta}(k+1) = \hat{\theta}(k) + F \frac{\phi(k)}{1 + \phi(k) F \phi^T(k)} \varepsilon_o(k)$$

with

$$\varepsilon_o(k) = y(k+1) - \hat{\theta}^T(k)\phi(k)$$

3. Recursive OLS

$$y(k+1) = \phi^T(k+1)\,\theta \quad = \quad [-y(k) \quad u(k)] \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix} y(1) \\ \vdots \\ y(k+1) \end{bmatrix} = \begin{bmatrix} \phi^T(1) \\ \vdots \\ \phi^T(k+1) \end{bmatrix} \theta$$

$$\rightarrow \hat{\theta} = (\phi^T\phi)^{-1}\phi^T y$$

$$\begin{matrix} y(k) \\ y(k+1) \end{matrix} = \begin{matrix} \phi(k) \\ \phi^T(k) \end{matrix} \quad \theta$$

$$D(t) = [\phi^T(k) \quad \phi(k)]^{-1}$$

$$\rightarrow \hat{\theta}(k+1) = \hat{\theta}(k) + D(k)\frac{\phi(k)}{1+\phi(k)D(k)\phi^T(k)}\varepsilon_0(k)$$

결론, Linear 문제 정의 등의 이유로.
$\rightarrow$ Same as gradient Search.

with $F(t) \longrightarrow D(t)$ $\quad$ by $D^{-1}(k+1) = D^{-1}(k) + \phi(k)\phi^T(k)$

Kalman $\approx$ $\langle \varepsilon \rangle$

$$\hat{\theta}(k+1) = \theta(k) + D(k)\frac{\phi(k)}{1+\phi(k)D(k)\phi^T(k)}\varepsilon_0(k)$$

INITIALIZE

$\hat{\theta}(1) = E[\theta]$

$D(1) = \mathrm{cov}(\theta)$

$\lfloor$ if different $\rightarrow$ bias $\hat{\theta}(k) \neq \theta$

with

$$\varepsilon_0(k) = y(k+1) - \hat{\theta}^T(k)\phi(k)$$

&

$$D^{-1}(k+1) = D^{-1}(k) + \phi(k)\phi^T(k)$$

- #1:

run OLS ~~on~~ a subset of data

- #2:

Take random value for θ + $\theta = [0, \dots, 0]$

Take "Large" value for D

- Interest of the RLS

> Computational Efficiency

invert only $N \times N$ and Not $N \times T$ (Large)

> Very adapted for ONLINE

4. Recursive OLS with Forgetting factor: λ

OLS minimizes

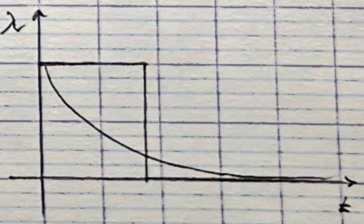
$$\mathcal{J}_k^{(n)} = \sum_{i=1}^{k} \varepsilon^2(i)$$

Minimize

$$\mathcal{J}_k^{(\sim)} = \sum_{i=1}^{k} \lambda_i \, \varepsilon^2(i)$$

λ: 0 — 1

과거로 갈수록 $\lambda \downarrow$



$$D^{-1}(k+1) = \lambda D^{-1}(k) + \phi(k)\,\phi^T(k)$$

5. <u>D(k) Update</u> (D)

$$D(k+1) = \frac{1}{\lambda}\left[D(k) - \frac{D(k)\,\phi^T \phi\, D(k)}{\lambda + \phi^T D(k)\,\phi}\right]$$

$$D = \frac{1}{\lambda}\left[D - \frac{D\phi^T\phi D}{\lambda + \phi^T D \phi}\right]$$

if $\lambda = 1 \Rightarrow D$

$\lambda \neq 1$

2023 NOV 22 Ex) MATLAB

$CD = \phi^T \theta$

$\phi = (1, \alpha)$

$\theta = (CD_\phi, CD_\alpha)$

$Y = PHI\ \theta$

$\hat{\theta} = (PHI^T PHI)^{-1} PHI^T Y$

$CD_\phi = \theta(1)$

$CD_\alpha = \theta(2)$

* Try other combi.

PHI_name = { ①, ②, ... more? }

\+ etc.

$$C_{m_{ac}} = C_{m_{ac,0}} + C_{m_{ac,\alpha}}\alpha$$

$+ C_{m_{ac,\beta}}\beta$ } → will probably have Large Variance

$+ \cdots$

$+ C_{m_{ac,\alpha^2}}\alpha^2$ } → 해볼 것

$$y(t) = \frac{b_1 z^{-1}}{1 + a_1 z^{-1}} u(t)$$

$$y(t) + a_1 y(t-1) = b_1 u(t-1)$$

$$\begin{aligned} y(t) &= -a_1 y(t-1) + b_1 u(t-1) \\ &= [-y(t-1)\ u(t)]\,[a_1,\ b_1]^T \\ &= \phi^T(t)\,\theta \end{aligned}$$

$Y = PHI\,\theta$ $\qquad \hat{\theta} = \sim$

<u>Recursive</u>

$$\hat{\theta}(t+1) = \hat{\theta}(t) + \frac{D(t)\phi(t)}{1 + \phi^T(t) D(t) \phi(t)} \Sigma_\circ(t+1)$$

■ How good is the model $\hat{\theta}$

- LS ~~is~~ is Optimal

as long as

$$y(t) = \phi^T(t)\,\theta + \varepsilon(t)$$

&

$$\begin{cases} E[\varepsilon(t)] = 0 \\ \varepsilon(t) \not\sim \phi(t) \end{cases}$$

Ex)

$u(t) \to [F] \to \otimes \to y(t)$, with ε entering at $\otimes$

$$y(t) = \frac{b_1 z^{-1}}{1 + a_1 z^{-1}} u(t) + \varepsilon(t)$$

$$y(t)\,[1 + a_1 z^{-1}] = b_1 z^{-1} u(t) + (1 + a_1 z^{-1})\,\varepsilon(t)$$

<A> $\varepsilon(t)$: White Noise

$$y(t) = (-y(t-1)\ u(t-1))\,(a_1, b_1)^T + \underbrace{\varepsilon(t) + a_1\varepsilon(t-1)}_{V(t)}$$

$V(t)$ is Correlated! with $\phi(t)$

$$y(t) = \phi(t)\,\theta^T + V(t)$$

ϕ^T (

$$\phi^T y(t) = \phi^T(t)\,\theta^T + \underbrace{\phi^T(t)\,V(t)}_{E[\phi^T(t)\,V(t)]}$$

$\Downarrow$

<A> 0 if LS est.

$$\begin{cases} \text{OK if } V(t) \not\sim \varepsilon(t) \\ \text{NOK if } V(t) = \varepsilon(t) + a_1\varepsilon(t-1) \end{cases}$$

<C> If $V(t)$ is NOT White Noise,

Then you have Bias in the estimate ($\hat{\theta} \neq \theta$)

In most cases, $\hat{\theta}$ is not good (?) → in real life NOT work well

> <§1> : LS is OK If $V(t)$ is Small

> <§2> : Use other <M>s

Ex) ELS (Extended Least Square)

LS : $Ay = Bu + \varepsilon$ ① ← ok if ε is White Noise

LS : $Ay = Bu + C\varepsilon$ ② → ELS

for

$$y(t) = [-y(t-1)\; u(t)]\,[a_1, b_1]^T + \varepsilon(t) + c_1\varepsilon(t-1)$$

$$= [-y(t),\; u(t-1),\; \varepsilon(t-1)]\,[a_1, b_1, c_1] + \varepsilon(t)$$

Replace by

$$\hat{\varepsilon}(t-1) = y(t-1) - \phi^T\theta = \varepsilon_0(t-1)$$

$y = \frac{B}{A}u + \frac{C}{D}\varepsilon$ ③ Numerical (NOT Explicit)

> Criteria for Model Validation

$$
\begin{array}{c}
y(t) \quad u(t) \qquad t : [1, N] \\
\searrow \swarrow \\
y,\ PHI \\
\downarrow \\
\hat{\theta} = PHI \setminus y \\
\downarrow \\
\hat{y} = PHI\ \hat{\theta} \\
\downarrow \\
e = y - \hat{y}
\end{array}
$$

We want to Minimize this! ←

To Check

[1] Magnitude of Error $= \sqrt{\dfrac{e^T e}{N}}$

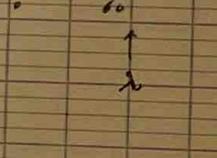
[2] e is White Noise → $E[e(t)] = 0$ & White Noise

[3] $e(t) \not\sim \phi(t)$ → 만족하기 힘든 조건

$e(t) \not\sim y(t)$

[4] $Cov(\theta)$

Task



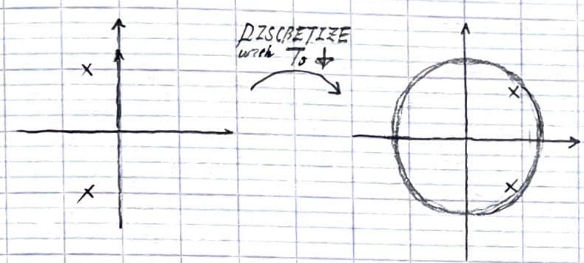
6. ID (G)

Ex) Code

C $F_1(s) = \dfrac{1}{1+0.1s+s^2}$

Equiv.

$T_s = 0.1$ [s] D $F_2(z)$



$\xrightarrow{u}$ [F] $\xrightarrow{y}$

$$y = \frac{b_1 s + b_0}{s^2 + a_1 s + a_0} u$$

$$y''(t) + a_1 y'(t) + a_0 y(t) = b_1 u'(t) + b_0 u(t)$$

$$y''(t) = [-y'(t), -y(t), u'(t), u(t)]\,[a_1, a_0, b_1, b_0]^T$$

<S> $t : [1, N]$

$$y'' = PHI \cdot \theta$$

⇒ θ is LS if we can

given $\{u(t), y(t)\}$

↓

Est. / Compute $\{y''(t), y'(t), u'(t)\}$

↓

θ obtained by LS → ⚠ Do NOT work in reality

<P> Est. of Derivatives. (FD with ↓h)

So →

> Classical 方: 상 Variable Filter (SVF)

$$s^2 y(t) + a_1 s y(t) + a_0 y(t) = b_1 s u(t) + b_0 u(t)$$

* Idea: pre-multiply by $\dfrac{1}{(s+\alpha)^2}$

"Filter Bandwidth"

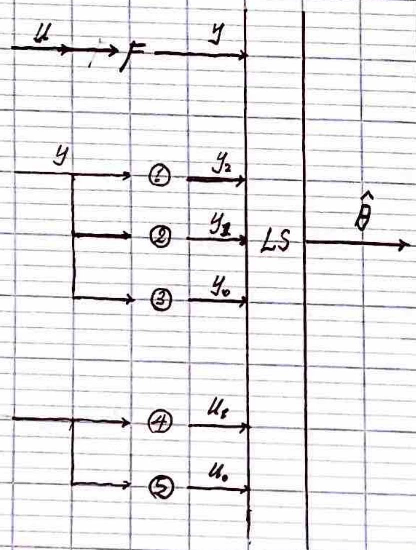
• α 너무 ↑:
→ Amplify Noise

• α 너무 ↓:
→ Lose significant signal.

$\times \dfrac{1}{(s+\alpha)^2}$

$$\overbrace{\underset{①}{\frac{s^2}{(s+\alpha)^2}} y(t)}^{y_2} + a_1 \overbrace{\underset{②}{\frac{s}{(s+\alpha)^2}} y(t)}^{y_1} + a_0 \overbrace{\underset{③}{\frac{1}{(s+\alpha)^2}} y(t)}^{y_0}$$

$$= \underbrace{b_1 \underset{④}{\frac{s}{(s+\alpha)^2}} u(t)}_{u_1} + \underbrace{b_0 \underset{⑤}{\frac{1}{(s+\alpha)^2}} u(t)}_{u_0}$$



"Pre-filtering"

Clean.

$$y_2 + a_1 y_1 + a_0 y_0 = b_1 u_1 + b_0 u_0$$

	↓ Freq.	↑ Freq.
①	1	0
②	s	0
③	s^2	1

In code
$\lambda = \alpha$

↳ does NOT work that well in Real life.

<S> it's ok like this

$$s^2 y(t) + a_1 s y(t) + a_0 y(t) = b_1 s u(t) + b_0 u(t) + \varepsilon(t)$$

$\frac{1}{(s+\alpha)^2}$ $\downarrow$ SVF

$$y_2(t) + a_1 y_1(t) + a_0 y_0(t) = b_1 u_1(t) + b_0 u_0(t) + \frac{1}{(s+\alpha)^2}\varepsilon(t)$$

NOT White Noise

<R> the Filtering

$\therefore \frac{1}{(s+\alpha)^2}\varepsilon(t) \propto \sim y_2(t)$

<C>
cannot run LS (?)

$$y_2 = \phi^T\theta + \varepsilon \qquad \phi = (-y_1, -y_0, u_1, u_0) \qquad \theta = (a_1, a_0, b_1, b_0)$$

<M1> LS $\phi^T y_2 = \phi^T\phi\theta^T + \phi^T\varepsilon$

<M2> Instrumental Variable §

IDEA: to find Z

$$Z^T y_2 = Z^T\phi\theta^T + Z^T\varepsilon$$

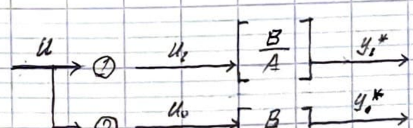
"Instrument" ← Z must be chosen §

- $Z^T\phi$: Invertible Ex) $Z=\phi$ ok.

★ • $Z \not\propto \varepsilon$ Ex) $Z=\phi$

§:

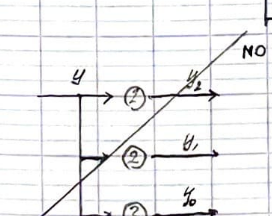
$$\phi = (-y_1, -y_0, u_1, u_0)$$



Then

$$Z = (-y_1^*, -y_0^*, u_1, u_0)$$

$$\hat{\theta} = (Z^T\phi)^{-1}Z^T y$$



$\hat{u}, y \rightarrow$ [SVF] $\rightarrow$ LS $\xrightarrow{A, B \text{ model}}$ NLS $\xrightarrow{A, B \text{ model}}$

<M>
Step resp. etc.

0 60 120

Coeff.
Δ

1. Take 1st 60 [s]

↓

Get the best 1D model.

- Validation <M>

2.

↓

6 model

- Validation <M>

3. Modify Recursive Algorithm
   ↳ Detect Change of Coeff.