Stochastic Adaptive Control (02421)

Lecture 10

Tobias K. S. Ritschel

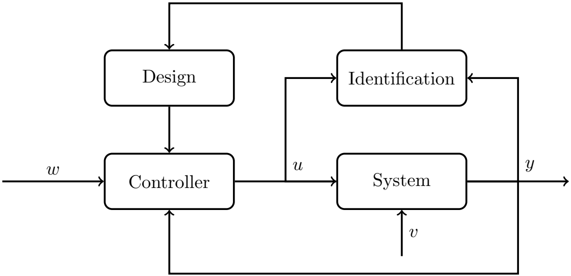
Section for Dynamical Systems

Department of Applied Mathematics and Computer Science

Technical University of Denmark

Lecture Plan

1. System theory
2. Stochastics
3. State estimation 1
4. State estimation 2
5. Optimal control 1
6. System identification 1 + adaptive control 1
7. External models + prediction
8. Optimal control 2
9. Optimal control 3
10. **System identification 2**
11. System identification 3 + model validation
12. System identification 4 + adaptive control 2
13. Adaptive control 3



Today's Agenda

- Least-squares (LS) estimation
- Maximum likelihood (ML) estimation
- Estimation for external models
- Recursive estimation

Follow-up from last time

Questions?

Estimation Methods

Observation equation

$$Y = G(\theta) + e, \quad y_t = g(t, \theta) + e_t \tag{1}$$

The noise e is zero-mean and has the variance $P = \sigma^2 \Sigma$

Residuals

$$\epsilon = Y - G(\hat{\theta}), \quad \epsilon_t = y_t - g(t, \hat{\theta}) \tag{2}$$

Linear case

$$G = \Phi\theta, \quad g(t, \theta) = \phi_t^T \theta \tag{3}$$

ϕ_t is a vector containing other data, such as inputs, past outputs, etc.

Least-squares estimation

Least Squares Method

Least squares (LS)

$$\min_{\theta} J_N(\theta) = \min_{\theta} \frac{1}{2} \sum_{t=1}^{N} \epsilon_t^2 = \min_{\theta} \frac{1}{2} \epsilon^T \epsilon \tag{4}$$

Solution

$$\left(\frac{\partial G(\theta)}{\partial \theta} \right)^T G(\theta) = \left(\frac{\partial G(\theta)}{\partial \theta} \right)^T Y \tag{5}$$

Linear case

$$\Phi^T \Phi \theta = \Phi^T Y, \quad \sum_{t=1}^{N} \phi_t \phi_t^T \theta = \sum_{t=1}^{N} \phi_t y_t \tag{6}$$

where Φ is

$$\Phi = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_N^T \end{bmatrix} \tag{7}$$

Least Squares Method

Parameter estimate

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y = \left(\sum_{t=1}^{N} \phi_t \phi_t^T \right)^{-1} \sum_{t=1}^{N} \phi_t y_t \tag{8}$$

$\Phi^T \Phi$ must have full rank

Distribution of estimate

$$\hat{\theta} \sim \mathcal{F}(\theta, P_\theta), \tag{9}$$

$$P_\theta = \mathrm{Cov}(\hat{\theta}) = (\Phi^T \Phi)^{-1} \Phi^T P \Phi (\Phi^T \Phi)^{-1} \tag{10}$$

Uncorrelated noise $(\Sigma = I)$

$$P_\theta = \sigma^2 (\Phi^T \Phi)^{-1} \tag{11}$$

Estimate of noise covariance (if it is unknown, but normally distributed)

$$\mathrm{Cov}(\hat{\theta}) \approx \hat{\sigma}^2 \left(\frac{\partial^2 J_N}{\partial \theta^2} (\hat{\theta}) \right)^{-1}, \qquad \hat{\sigma}^2 \approx 2 \frac{J_N(\hat{\theta})}{N - n_\theta} \tag{12}$$

Least Squares Method – Main properties

Properties of linear least squares estimators

- It is a linear function of the observations, Y
- It is unbiased: $\mathbb{E}[\hat{\theta}] = \theta$ and $\text{Cov}(\hat{\theta}) = (\Phi^T\Phi)^{-1}\Phi^T P\Phi(\Phi^T\Phi)^{-1}$
- It does not assume a specific distribution

If $P = \sigma^2 I$

- Unbiased: $\mathbb{E}[\hat{\theta}] = \theta$ and $\text{Cov}(\hat{\theta}) = \sigma^2(\Phi^T\Phi)^{-1}$
- Independent: $\epsilon \perp \hat{\theta}$
- $\hat{\theta}$ is the best linear unbiased estimator (BLUE), which means that it has the smallest variance among all estimators which are linear functions of the observations

Estimation

Linear parametrized model

$$y_t = a + bu_t + cu_{t-1}^2 + e_t = \begin{bmatrix} 1 & u_t & u_{t-1}^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} + e_t \tag{13}$$

$$= \phi^T \theta + e_t \tag{14}$$

Unbiased estimator

$$\mathbb{E}[\hat{\theta} - \theta] = 0 \tag{15}$$

Minimal variance estimator

$$\mathrm{Cov}(\hat{\theta} - \theta | \theta) \leq \mathrm{Cov}(\bar{\theta} - \theta | \theta) \tag{16}$$

for any other estimator $\bar{\theta}$

Linear Models

If we consider the parameters

$$\theta^T = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix} \tag{17}$$

which of the following models are linear in the sense of estimation?

- $y_t = \theta_1 u_t + e_t$
- $y_t = \theta_1 u_t + \theta_2 u_t x_t + e_t$
- $y_t = \theta_1 u_t + \theta_2 \theta_3 x_t + e_t$
- $y_t = cos(\theta_1) u_t + \theta_2 z_t + \theta_3 x_t + e_t$
- $y_t = cos(\theta_1) u_t + \theta_2 z_t + \theta_3 \theta_1 x_t + \theta_1 y_{t-1} + e_t$
- $y_t = cos(\theta_1 u_t) + \theta_2 z_t + \theta_3 \theta_1 x_t + \theta_1 y_{t-1} + e_t$

Think about it for yourself for <u>one minute</u> and
then discuss with the person next to you for <u>one minute</u>.

Least squares – Example

System

$$y_t = \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + e_t \tag{18}$$

$$= \underbrace{\begin{bmatrix} u_t & u_{t-1} & u_{t-2} \end{bmatrix}}_{\phi_t^T} \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}}_{\theta} + e_t \tag{19}$$

Matrix ($N = 3$ measurements)

$$\Phi = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \phi_3^T \end{bmatrix} = \begin{bmatrix} u_1 & u_0 & u_{-1} \\ u_2 & u_1 & u_0 \\ u_3 & u_2 & u_1 \end{bmatrix} \tag{20}$$

Least squares – Example

Measurements

$$y_1 = 1, \qquad y_2 = 2, \qquad y_3 = 3 \tag{21}$$

Inputs

$$u_{-1} = 3, \qquad u_0 = 1, \qquad u_1 = 4, \qquad u_2 = -1, \qquad u_3 = 2 \tag{22}$$

Matrix

$$\Phi = \begin{bmatrix} 4 & 1 & 3 \\ -1 & 4 & 1 \\ 2 & -1 & 4 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{23}$$

Parameter estimate

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = (\Phi^T\Phi)^{-1}\Phi^T Y = \begin{bmatrix} -0.5660 \\ 0.0943 \\ 1.0566 \end{bmatrix} \tag{24}$$

Least squares – Exercise

Measurements

$$y_1 = 4, \qquad y_2 = -1, \qquad y_3 = 2 \tag{25}$$

Inputs

$$u_{-1} = 5, \qquad u_0 = 2, \qquad u_1 = -2, \qquad u_2 = -3, \qquad u_3 = 1 \tag{26}$$

Solve the exercise in 10 min.

Matrix

$$\Phi = \begin{bmatrix} -2 & 2 & 5 \\ -3 & -2 & 2 \\ 1 & -3 & -2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix} \tag{27}$$

Parameter estimate

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = (\Phi^T \Phi)^{-1} \Phi^T Y = \begin{bmatrix} 2.9259 \\ -1.3704 \\ 2.5185 \end{bmatrix} \tag{28}$$

Maximum likelihood estimation

Maximum Likelihood Method

Likelihood

$$\mathcal{L}(\theta) = f(Y|\theta) \tag{29}$$

Maximum likelihood estimation problem (equivalent formulations)

$$\max_{\theta} \mathcal{L}(\theta), \qquad \max_{\theta} \ln \mathcal{L}(\theta) \tag{30}$$

Maximum likelihood estimation requires an assumption of the distribution

Maximum Likelihood Method

Assume that $Y = \Phi\theta + e$ and that $e \sim N(0, P)$

Probability distribution of N observations

$$f(Y|\theta) = \frac{1}{\sqrt{(2\pi)^N \sqrt{\det P}}} \exp\left(-\frac{1}{2}(Y - \Phi\theta)^T P^{-1}(Y - \Phi\theta)\right) \quad (31)$$

Log-likelihood function

$$\ln \mathcal{L}(Y;\theta) = -\frac{1}{2}\ln\det P - \frac{N}{2}\ln 2\pi - \frac{1}{2}(Y - \Phi\theta)^T P^{-1}(Y - \Phi\theta) \quad (32)$$

Optimization problem

$$\max_\theta \ \ln(\mathcal{L}(\theta)) = \min_\theta \ -\ln\mathcal{L}(\theta) \quad (33)$$

$$= \min_\theta \frac{1}{2}\ln\det P + \frac{1}{2}(Y - \Phi\theta)^T P^{-1}(Y - \Phi\theta) + c \quad (34)$$

c is a constant independent of θ and P

Maximum Likelihood Method

First-order optimality conditions

$$\frac{\partial \ln \mathcal{L}}{\partial \theta}(Y;\theta) = \frac{1}{2}(-2\Phi^T P^{-1} Y + 2\Phi^T P^{-1}\Phi\theta) = 0 \tag{35}$$

Optimal estimate

$$\hat{\theta} = (\Phi^T P^{-1}\Phi)^{-1}\Phi^T P^{-1} Y \tag{36}$$

Only the structure Σ of the variance $P = \sigma^2\Sigma$ is important

$$\hat{\theta} = (\Phi^T \Sigma^{-1}\Phi)^{-1}\Phi^T \Sigma^{-1} Y \tag{37}$$

If $(P = \sigma^2 I)$, the MLE estimator is identical with the LS estimator

$$\hat{\theta} = \frac{\sigma^2}{\sigma^2}(\Phi^T\Phi)^{-1}\Phi^T Y = (\Phi^T\Phi)^{-1}\Phi^T Y \tag{38}$$

Maximum Likelihood Method

ML is based on the assumption that Σ is known, but σ^2 can be unknown

First-order optimality conditions for σ^2

$$\frac{\partial \ln \mathcal{L}}{\partial \sigma^2}(Y;\theta) = \frac{N}{2\sigma^2} - \frac{1}{2\sigma^4}(Y - \Phi\theta)^T \Sigma^{-1}(Y - \Phi\theta) = 0 \tag{39}$$

$\det P = (\sigma^2)^N \det \Sigma$

ML estimate of the noise covariance

$$\hat{\sigma}^2 = \frac{(Y - \Phi\hat{\theta})^T \Sigma^{-1}(Y - \Phi\hat{\theta})}{N} \tag{40}$$

Maximum Likelihood Method – Main properties

Properties of the ML estimator (assuming a normal distribution)

- It is unbiased: $\hat{\theta} \sim N\{\theta, (\Phi^T \Sigma^{-1} \Phi)^{-1} \Phi^T \Sigma^{-1} P \Sigma^{-1} \Phi (\Phi^T \Sigma^{-1} \Phi)^{-1}\}$
- It is a linear function of the observations, Y

and for the case $P = \sigma^2 I$

- The estimate is equivalent to the LS estimator
- It is unbiased: $\hat{\theta} \sim N\{\theta, \sigma^2 (\Phi^T \Phi)^{-1}\}$
- Independent: $\epsilon \perp \hat{\theta}$
- $\hat{\theta}$ is the best linear unbiased estimator (BLUE), which means that it has the smallest variance among all estimators which are linear functions of the observations

Residual-Estimator Independence

Both the LS and ML estimators achieve residual-estimator independence when $P = \sigma^2 I$

Covariance

$$\begin{aligned}
\operatorname{Cov}(\epsilon, \hat{\theta}) &= \operatorname{Cov}(Y - \Phi\hat{\theta}, \hat{\theta}) && (41)\\
&= \operatorname{Cov}(\Phi\theta + e - \Phi\hat{\theta}, \hat{\theta}), \quad Y = \Phi\theta + e && (42)\\
&= \operatorname{Cov}(e, \hat{\theta}) - \Phi \operatorname{Cov}(\hat{\theta}, \hat{\theta}) && (43)\\
&= \operatorname{Cov}(e, e)L^T - \Phi L \operatorname{Cov}(e, e)L^T, \quad \hat{\theta} = LY = L\Phi\theta + Le && (44)\\
&= (I - \Phi L)PL^T && (45)
\end{aligned}$$

LS estimator ($L = (\Phi^T\Phi)^{-1}\Phi^T$)

$$\operatorname{Cov}(\epsilon, \hat{\theta}) = (I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)P\Phi(\Phi^T\Phi)^{-1} \quad (46)$$

If $P = \sigma^2 I$ is a multiple of the identity matrix

$$\begin{aligned}
\operatorname{Cov}(\epsilon, \hat{\theta}) &= \sigma^2(I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)\Phi(\Phi^T\Phi)^{-1} && (47)\\
&= \sigma^2(\Phi(\Phi^T\Phi)^{-1} - \Phi(\Phi^T\Phi)^{-1}\Phi^T\Phi(\Phi^T\Phi)^{-1}) = 0 && (48)
\end{aligned}$$

Maximum likelihood – Example

Measurements

$$y_1 = 1, \qquad y_2 = 2, \qquad y_3 = 3, \qquad P = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix} \tag{49}$$

Inputs

$$u_{-1} = 3, \qquad u_0 = 1, \qquad u_1 = 4, \qquad u_2 = -1, \qquad u_3 = 2 \tag{50}$$

Matrix

$$\Phi = \begin{bmatrix} 4 & 1 & 3 \\ -1 & 4 & 1 \\ 2 & -1 & 4 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{51}$$

Parameter estimate

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = (\Phi^T P^{-1} \Phi)^{-1} \Phi^T P^{-1} Y = \begin{bmatrix} -0.5660 \\ 0.0943 \\ 1.0566 \end{bmatrix} \tag{52}$$

Maximum likelihood – Exercise

Measurements and noise covariance

$$y_1 = 4, \qquad y_2 = -1, \qquad y_3 = 2, \qquad P = \begin{bmatrix} 2 & & \\ & 4 & \\ & & 5 \end{bmatrix} \tag{53}$$

Inputs

$$u_{-1} = 5, \qquad u_0 = 2, \qquad u_1 = -2, \qquad u_2 = -3, \qquad u_3 = 1 \tag{54}$$

Solve the exercise in 10 min.

Matrix

$$\Phi = \begin{bmatrix} -2 & 2 & 5 \\ -3 & -2 & 2 \\ 1 & -3 & -2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix} \tag{55}$$

Parameter estimate

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = (\Phi^T P^{-1} \Phi)^{-1} \Phi^T P^{-1} Y = \begin{bmatrix} 2.9259 \\ -1.3704 \\ 2.5185 \end{bmatrix} \tag{56}$$

Identification of external models

Identification of external models

Next, we will apply the estimation methods to different models

- ARX: $A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t$
- OE: $y_t = q^{-k}\frac{B(q^{-1})}{A(q^{-1})}u_t + e_t$
- IV: $A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t$, where e_t is not white
- ARMAX: $A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t$

ARX estimation

ARX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t, \quad (57)$$

$$y_t = -\sum_{i=1}^{n_a} a_i y_{t-i} + \sum_{i=0}^{n_b} b_i u_{t-i-k} + e_t \quad (58)$$

$e_t \sim F(0, P)$ and white

Rewrite

$$y_t = \sum_{i=1}^{n_\phi} \theta_i \phi_{t,i} + e_t = \phi_t^T \theta + e_t, \quad (59)$$

$$\phi_t^T = [-y_{t-1}, -y_{t-2}, ..., -y_{t-n_a}, u_{t-k}, ..., u_{t-k-n_b}], \quad (60)$$

$$\theta^T = [a_1, a_2, ..., a_{n_a}, b_0, b_1, ..., b_{n_b}] \quad (61)$$

Estimation - LS and ML

Least-squares method

$$Y_t = \Phi_t\theta + E_t, \quad E_t \sim \mathbb{F}(0, P) \tag{62}$$

$$\hat{\theta} = (\Phi_t^T\Phi_t)^{-1}\Phi_t^T Y_t \tag{63}$$

$$\Rightarrow \hat{\theta} \sim \mathbb{F}(\theta, (\Phi_t^T\Phi_t)^{-1}\Phi_t^T P\Phi_t(\Phi_t^T\Phi_t)^{-1}) \tag{64}$$

Maximum-likelihood method

$$Y_t = \Phi_t\theta + E_t, \quad E_t \sim N(0, P) \tag{65}$$

$$\hat{\theta} = (\Phi_t^T P^{-1}\Phi_t)^{-1}\Phi_t^T P^{-1} Y_t \tag{66}$$

$$\Rightarrow \hat{\theta} \sim \mathbb{F}(\theta, (\Phi_t^T P^{-1}\Phi_t)^{-1}\Phi_t^T P^{-1}\Phi_t(\Phi_t^T P^{-1}\Phi_t)^{-1}) \tag{67}$$

If $P = \sigma^2\Sigma$

$$\hat{\sigma}^2 = \frac{(Y - \Phi_t\hat{\theta})^T\Sigma^{-1}(Y - \Phi_t\hat{\theta})}{N} \tag{68}$$

ARX example

Model

$$y_t = \frac{-1}{4} y_{t-1} + \frac{1}{2} y_{t-2} + u_{t-1} + e_t \quad (69)$$

u_t is an integer random input sequence that excites the system. 10 samples and the system variance is $\sigma^2 = 0.1$

$$y_{-2:9} = \begin{bmatrix} 0 & 1 & 0.76 & 1.32 & 1.05 & 1.34 & 1.15 & 1.42 & 2.20 & 2.15 & 2.57 & 2.42 \end{bmatrix}$$

$$u_{-1:9} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

Compute Φ and use that $\theta = (\Phi_t^T \Phi_t)^{-1} \Phi_t^T Y_t$ and $P_\theta = (\Phi_t^T \Phi_t)^{-1} \Phi_t^T P \Phi_t (\Phi_t^T \Phi_t)^{-1}$

$$\theta = \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix}, \qquad P_\theta = \begin{bmatrix} 0.1355 & -0.0431 & -0.1057 \\ -0.0431 & 0.0620 & -0.0088 \\ -0.1057 & -0.0088 & 0.1242 \end{bmatrix}$$

How would you verify that the covariance matrix is correct?
Think about it for yourself for one minute and
then discuss with the person next to you for one minute.

ARX exercise

Model

$$y_t = \frac{1}{3}y_{t-1} - \frac{1}{3}y_{t-2} + 2u_{t-1} + e_t \tag{70}$$

u_t is an integer random input sequence that excites the system

10 samples and the system variance is $\sigma^2 = 0.1$

$$y_{-2:9} = \begin{bmatrix} 1 & 1 & 2.17 & 2.97 & 1.55 & 1.80 & 4.18 & 2.38 & 3.26 & 2.40 & 2.84 & 5.02 \end{bmatrix}$$

$$u_{-1:9} = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 1 & 2 & 1 & 1 & 2 & 2 \end{bmatrix}$$

Solve the exercise in 15 min.

Compute Φ and use that $\theta = (\Phi_t^T \Phi_t)^{-1} \Phi_t^T Y_t$ and
$P_\theta = (\Phi_t^T \Phi_t)^{-1} \Phi_t^T P \Phi_t (\Phi_t^T \Phi_t)^{-1}$

$$\theta = \begin{bmatrix} 0.2170 \\ -0.2406 \\ 2.1909 \end{bmatrix}, \qquad P_\theta = \begin{bmatrix} 0.0081 & -0.0043 & -0.0065 \\ -0.0043 & 0.0117 & -0.0118 \\ -0.0065 & -0.0118 & 0.0353 \end{bmatrix}$$

Estimation - OE methods

Input/output relation

$$y_t = q^{-k}\frac{B(q^{-1})}{A(q^{-1})}u_t + e_t \tag{71}$$

Estimate is based on minimization of the error

$$J = \frac{1}{2}\sum_{t=1}^{N}\epsilon_t^2 = \frac{1}{2}\sum_{t=1}^{N}\left(y_t - q^{-k}\frac{B(q^{-1})}{A(q^{-1})}u_t\right)^2 \tag{72}$$

Notice the difference to the LS where $\epsilon_t = A(q^{-1})y_t - q^{-k}B(q^{-1})u_t$

- An OE method is more resistant towards higher sampling frequencies than 1-step prediction methods
- The OE methods require output and input data and the polynomial orders
- One of the disadvantages of OE methods is that the cost is not quadratic in the parameters

Estimation - IV-method (instrumental variable)

IV-method

$$\hat{\theta} = \left(\sum_{t=1}^{N} \psi_t \phi_t \right)^{-1} \sum_{t=1}^{N} \psi_t^T y_t = \left(\Psi \Phi \right)^{-1} \Psi^T Y \tag{73}$$

$\psi_t \approx \phi_t$ is chosen such that

$$E\{\psi_t e_t\} = 0 \tag{74}$$

$$E\{\psi_t \phi_t^T\} \text{ is invertible} \tag{75}$$

Example

$$\phi_t = [-y_{t-1}, -y_{t-2}, ..., -y_{t-n_a}, u_t, u_{t-k}, ..., u_{t-k-n_b}] \tag{76}$$

$$\psi_t = [-\bar{y}_{t-1}, -\bar{y}_{t-2}, ..., -\bar{y}_{t-n_a}, u_t, u_{t-k}, ..., u_{t-k-n_b}] \tag{77}$$

$\bar{y}_t$ is an estimated output, $\bar{y}_t = H_{est}(q)u_t$, e.g., based on previous parameter estimates

Estimation - Bayes estimation

ARX model

$$A(q^{-1})y_t = B(q^{-1})u_t + e_t, \quad e_t \perp e_s, \quad e_t \sim N(0, \sigma_t^2) \tag{78}$$

Model at time t and $t+1$

$$y_t = \phi_t^T \theta + e_t \tag{79}$$

$$y_{t+1} = \phi_{t+1}^T \theta + e_{t+1} \tag{80}$$

Bayesian estimation (conditional estimator)

$$\theta | Y_t \sim N(\hat{\theta}_t, P_t) \tag{81}$$

Note: It is similar to estimation with the Kalman filter

Estimation - Bayes estimation

Estimator

$$\hat{\theta}_{t+1} = \hat{\theta}_t + K_{t+1}(y_{t+1} - \phi_{t+1}^T \hat{\theta}_t) \tag{82}$$

$$K_{t+1} = \frac{P_{t+1}\phi_{t+1}}{\sigma_{t+1}^2} = P_t \phi_{t+1}(\phi_{t+1}^T P_t \phi_{t+1} + \sigma_{t+1}^2)^{-1} \tag{83}$$

$$P_{t+1}^{-1} = P_t^{-1} + \frac{\phi_{t+1}\phi_{t+1}^T}{\sigma_{t+1}^2}, \quad P_{t+1} = (I - K_{t+1}\phi_{t+1}^T)P_t \tag{84}$$

Estimation error

$$\tilde{\theta}|Y_t \sim N(0, P_t) \tag{85}$$

Approximate error distribution (due to the asymptotic behaviour of the estimator)

$$\tilde{\theta}_t \sim N(0, P_t) \tag{86}$$

ARMAX estimation

ARMAX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t \quad (87)$$

Rewrite (same form as ARX model)

$$y_t = \phi_t^T\theta + e_t \quad (88)$$

$$\phi_t = [-y_{t-1}, -y_{t-2}, ..., -y_{t-n_a}, u_{t-k}, u_{t-k-1}, ..., u_{t-k-n_b}, e_{t-1}, e_{t-2}, ..., e_{t-n_c}]^T \quad (89)$$

N consecutive measurements

$$Y_N = \Phi_N\theta + E_N \quad (90)$$

$$Y_N = [y_1, ..., y_N]^T \qquad E_N = [e_1, ..., e_N]^T \quad (91)$$

Φ_N includes estimates of e

ARMAX estimation - Extended Least-Squares

Approximate noise by residuals

$$e_t \approx \epsilon_t = y_t - \phi_t^T \hat{\theta} \tag{92}$$

ϕ_t includes previous noise estimates

Rewrite as transfer functions

$$\epsilon_t = \frac{\hat{A}(q^{-1})}{\hat{C}(q^{-1})} y_t - \frac{\hat{B}(q^{-1})}{\hat{C}(q^{-1})} u_t \tag{93}$$

Estimator and error variance (note that n is the iteration index - not time)

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \left(\sum_{i=1}^{N} \phi_i \phi_i^T \right)^{-1} \sum_{i=1}^{N} \phi_i \epsilon_i \tag{94}$$

$$P_N = \frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2 \times \left(\sum_{i=1}^{N} \phi_i \phi_i^T \right)^{-1} \tag{95}$$

ARMAX estimation - Maximum Likelihood

ARMAX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t \tag{96}$$

Residuals

$$\epsilon_t = \frac{\hat{A}(q^{-1})}{\hat{C}(q^{-1})}y_t - \frac{\hat{B}(q^{-1})}{\hat{C}(q^{-1})}u_t = y_t - \phi_t^T\hat{\theta} \tag{97}$$

In ML, we consider the noise $e_t \sim N(0, \sigma^2)$.

ARMAX estimation - Maximum Likelihood

Design estimator using Newton's iterative method (note that n is the iteration index - not time)

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \left(\sum_{i=1}^{N} \psi_i \psi_i^T \right)^{-1} \sum_{i=1}^{N} \psi_i \epsilon_i \tag{98}$$

ψ_t is a filtered version of ϕ_t

$$\psi_t = \frac{1}{\hat{C}(q^{-1})} \phi_t \tag{99}$$

It can be computed through a state-space realization or using past values

$$\psi_t = \phi_t - \sum_{i=1}^{n_c} \hat{c}_i \psi_{t-i} \tag{100}$$

Commands - Matlab

Useful Matlab commands:

1. arx(Data, [orders]) - least-Squares methods
2. oe(Data, [orders]) - output error methods
3. tfest(Data, [orders]) - transfer function estimation methods (OE)
4. iv4(Data, [orders]) - instrumental variable methods
5. bj(Data, [orders]) - least-squares method
6. armax(Data, [orders]) - least-squares method
7. pem(Data, [orders]) - prediction error estimation
8. polyest(Data, [orders]) - estimation for the L-structure
9. polydata(sys) - polynomial coefficients and uncertainties
10. getpvec(sys) - get vector of parameters
11. getcov(sys) - get parameter covariance matrix

Recursive parameter estimation

Recursive Estimation

The previously presented methods are in the form

$$\hat{\theta}_t = \mathrm{func}(Y_t) \tag{101}$$

We use all measurements up to and including time t, which becomes computationally intensive over time

Recursive methods only rely on the current measurement and the past estimate

$$\hat{\theta}_t = \mathrm{func}(y_t, \hat{\theta}_{t-1}) \tag{102}$$

- It assumes that $\hat{\theta}_{t-1}$ is a sufficient statistic of Y_{t-1}
- It can easily be adapted to account for time-varying parameters

Estimation - Estimation by Iteration

Least squares estimator

$$\hat{\theta}_t = (\Phi^T \Phi)^{-1} \Phi^T Y_t \tag{103}$$

If $Y_t = \Phi \bar{\theta} + \epsilon$ for some previous estimator $\bar{\theta}$

$$\hat{\theta}_t = \bar{\theta} + (\Phi^T \Phi)^{-1} \Phi^T \epsilon \tag{104}$$

Iterative formulation of the LS estimator

$$\hat{\theta}_t = \hat{\theta}_{t-1} + (\Phi_t^T \Phi_t)^{-1} \Phi_t^T \epsilon_t \tag{105}$$

RLS/RARX - Recursive Least Squares

ARX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t, \quad e_t \sim \mathcal{F}(0,\sigma^2) \tag{106}$$

$$y_t = \phi_t^T\theta + e_t, \quad e_t \perp e_s \quad s > t \tag{107}$$

$$\phi_t = [-y_{t-1},\dots,-y_{t-n_a},u_{t-k},\dots,u_{t-n_b-k}]^T \tag{108}$$

$$\theta = [a_1,\dots,a_{n_a},b_0,\dots,b_{n_b}]^T \tag{109}$$

Least squares estimator based on t measurements

$$\hat{\theta}_t = \left(\sum_{i=1}^{t}\phi_i\phi_i^T\right)^{-1}\sum_{i=1}^{t}\phi_i y_i, \tag{110}$$

$$P_t^{-1} = \sum_{i=1}^{t}\phi_i\phi_i^T, \quad \sum_{i=1}^{t}\phi_i\epsilon_i = 0 \tag{111}$$

RLS/RARX - Recursive Least Squares

Recursive formulation

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \sum_{i=1}^{t} \phi_i \epsilon_i \tag{112}$$

Rewrite the recursion

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \phi_t \epsilon_t \tag{113}$$

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \tag{114}$$

$$P_t^{-1} = P_{t-1}^{-1} + \phi_t \phi_t^T \tag{115}$$

$$\mathrm{Var}(\hat{\theta}_t | Y_t) = P_t \sigma^2 \approx \mathrm{Var}(\hat{\theta}_t) \tag{116}$$

If no a priori knowledge about the parameter values is available, use

$$\hat{\theta}_0 = 0, \quad P_0 = \beta I, \quad \beta \gg 0 \tag{117}$$

RLS/RARX - Recursive Least Squares

The recursion can also be computed using alternative formulations

Example (inspired by the Hemes' inversion lemma and square-root/factorization algorithms)

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \tag{118}$$

$$s_t = 1 + \phi_t^T P_{t-1} \phi_t \tag{119}$$

$$K_t = \frac{P_{t-1}\phi_t}{s_t} \tag{120}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t \epsilon_t \tag{121}$$

$$P_t = P_{t-1} - K_t s_t K_t^T \tag{122}$$

RLS/RARX - Example

Model

$$y_t = \frac{-1}{4} y_{t-1} + \frac{1}{2} y_{t-2} + u_{t-1} + e_t \tag{123}$$

New measurement ($t = 10$)

$$y_t = 1.47, \qquad u_{t-1} = 2 \tag{124}$$

Covariates, parameter estimate and covariance

$$\phi_t = \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix}, \quad \hat{\theta}_{t-1} = \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix}, \quad P_{t-1} = \begin{bmatrix} 0.1355 & -0.0431 & -0.1057 \\ -0.0431 & 0.0620 & -0.0088 \\ -0.1057 & -0.0088 & 0.1242 \end{bmatrix}$$

RLS/RARX - Example

Innovation error and auxiliary variable

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} = 1.47 - \begin{bmatrix} 2.42 & 2.57 & 2 \end{bmatrix} \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix} = -1.20, \tag{125}$$

$$s_t = 1 + \phi_t^T P_{t-1} \phi_t \tag{126}$$

$$= 1 + \begin{bmatrix} 2.42 & 2.57 & 2 \end{bmatrix} \begin{bmatrix} 0.1355 & -0.0431 & -0.1057 \\ -0.0431 & 0.0620 & -0.0088 \\ -0.1057 & -0.0088 & 0.1242 \end{bmatrix} \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix} = 1.05 \tag{127}$$

Gain

$$K_t = \frac{P_{t-1}\phi_t}{s_t} = \frac{1}{1.05} \begin{bmatrix} 0.1355 & -0.0431 & -0.1057 \\ -0.0431 & 0.0620 & -0.0088 \\ -0.1057 & -0.0088 & 0.1242 \end{bmatrix} \begin{bmatrix} 2.42 \\ 2.57 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.0055 \\ 0.0357 \\ -0.0286 \end{bmatrix} \tag{128}$$

RLS/RARX - Example

Parameter estimate

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t \epsilon_t = \begin{bmatrix} -0.2505 \\ 0.4960 \\ 0.9991 \end{bmatrix} - 1.20 \begin{bmatrix} 0.0055 \\ 0.0357 \\ -0.0286 \end{bmatrix} = \begin{bmatrix} -0.2570 \\ 0.4533 \\ 1.0333 \end{bmatrix} \tag{129}$$

Covariance

$$P_t = P_{t-1} - K_t s_t K_t^T = \begin{bmatrix} 0.1355 & -0.0431 & -0.1057 \\ -0.0431 & 0.0620 & -0.0088 \\ -0.1057 & -0.0088 & 0.1242 \end{bmatrix} \tag{130}$$

$$- 1.05 \begin{bmatrix} 0.0055 \\ 0.0357 \\ -0.0286 \end{bmatrix} \begin{bmatrix} 0.0055 & 0.0357 & -0.0286 \end{bmatrix} \tag{131}$$

$$= \begin{bmatrix} 0.1355 & -0.0433 & -0.1055 \\ -0.0433 & 0.0607 & -0.0077 \\ -0.1055 & -0.0077 & 0.1233 \end{bmatrix} \tag{132}$$

RLS/RARX - Exercise

Model

$$y_t = \frac{1}{3}y_{t-1} - \frac{1}{3}y_{t-2} + 2u_{t-1} + e_t \quad (133)$$

New measurement ($t = 10$)

$$y_t = 4.30, \qquad u_{t-1} = 2 \quad (134)$$

Covariates, parameter estimate and covariance

$$\phi_t = \begin{bmatrix} 5.02 \\ 2.84 \\ 2 \end{bmatrix}, \quad \hat{\theta}_{t-1} = \begin{bmatrix} 0.2170 \\ -0.2406 \\ 2.1909 \end{bmatrix}, \quad P_{t-1} = \begin{bmatrix} 0.0081 & -0.0043 & -0.0065 \\ -0.0043 & 0.0117 & -0.0118 \\ -0.0065 & -0.0118 & 0.0353 \end{bmatrix}$$

Solve the exercise in 15 min.

RLS/RARX - Exercise

Innovation error and auxiliary variable

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} = 4.30 - \begin{bmatrix} 5.02 & 2.84 & 2 \end{bmatrix} \begin{bmatrix} 0.2170 \\ -0.2406 \\ 2.1909 \end{bmatrix} = -0.4878, \quad (135)$$

$$s_t = 1 + \phi_t^T P_{t-1} \phi_t \quad (136)$$

$$= 1 + \begin{bmatrix} 5.02 & 2.84 & 2 \end{bmatrix} \begin{bmatrix} 0.0081 & -0.0043 & -0.0065 \\ -0.0043 & 0.0117 & -0.0118 \\ -0.0065 & -0.0118 & 0.0353 \end{bmatrix} \begin{bmatrix} 5.02 \\ 2.84 \\ 2 \end{bmatrix} = 1.05 \quad (137)$$

Gain

$$K_t = \frac{P_{t-1}\phi_t}{s_t} = \frac{1}{1.05} \begin{bmatrix} 0.0081 & -0.0043 & -0.0065 \\ -0.0043 & 0.0117 & -0.0118 \\ -0.0065 & -0.0118 & 0.0353 \end{bmatrix} \begin{bmatrix} 5.02 \\ 2.84 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.0147 \\ -0.0114 \\ 0.0042 \end{bmatrix} \quad (138)$$

RLS/RARX - Exercise

Parameter estimate

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t\epsilon_t = \begin{bmatrix} 0.2170 \\ -0.2406 \\ 2.1909 \end{bmatrix} - 0.49 \begin{bmatrix} 0.0147 \\ -0.0114 \\ 0.0042 \end{bmatrix} = \begin{bmatrix} 0.2098 \\ -0.2351 \\ 2.1888 \end{bmatrix} \quad (139)$$

Covariance

$$P_t = P_{t-1} - K_t s_t K_t^T = \begin{bmatrix} 0.0081 & -0.0043 & -0.0065 \\ -0.0043 & 0.0117 & -0.0118 \\ -0.0065 & -0.0118 & 0.0353 \end{bmatrix} \quad (140)$$

$$- 1.05 \begin{bmatrix} 0.0147 \\ -0.0114 \\ 0.0042 \end{bmatrix} \begin{bmatrix} 0.0147 & -0.0114 & 0.0042 \end{bmatrix} \quad (141)$$

$$= \begin{bmatrix} 0.0079 & -0.0041 & -0.0066 \\ -0.0041 & 0.0116 & -0.0117 \\ -0.0066 & -0.0117 & 0.0353 \end{bmatrix} \quad (142)$$

RELS - Recursive Extended Least Squares

ARMAX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t \tag{143}$$

$$y_t = \phi_t^T\theta + e_t \tag{144}$$

$$\phi_t = [-y_{t-1}, \ldots, -y_{t-n_a}, u_{t-k}, \ldots, u_{t-n_b-k}, e_{t-1}, \ldots, e_{t-n_c}]^T \tag{145}$$

$$\theta = [a_1, \ldots, a_{n_a}, b_0, \ldots, b_{n_b}, c_1, \ldots, c_{n_c}]^T \tag{146}$$

Least squares estimator (approximate e_i by ϵ_i in ϕ)

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t\phi_t\epsilon_t \tag{147}$$

$$\epsilon_i = y_i - \phi_i^T\hat{\theta}_{i-1} \tag{148}$$

$$P_t^{-1} = P_{t-1}^{-1} + \phi_t\phi_t^T \tag{149}$$

RML - Recursive Maximum Likelihood

ARMAX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t \tag{150}$$

$$y_t = \phi_t^T\theta + e_t \tag{151}$$

$$\phi_t = [-y_{t-1}, \ldots, -y_{t-n_a}, u_{t-k}, \ldots, u_{t-n_b-k}, e_{t-1}, \ldots, e_{t-n_c}]^T \tag{152}$$

$$\theta = [a_1, \ldots, a_{n_a}, b_0, \ldots, b_{n_b}, c_1, \ldots, c_{n_c}]^T \tag{153}$$

Maximum likelihood estimator (approximate e_i by ϵ_i in ϕ)

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t\psi_t\epsilon_t, \quad \psi_t = \frac{1}{\hat{C}(q^{-1})}\phi_t \tag{154}$$

$$\epsilon_i = y_i - \phi_i^T\hat{\theta}_{i-1} \tag{155}$$

$$P_t^{-1} = P_{t-1}^{-1} + \psi_t\psi_t^T \tag{156}$$

Questions

Questions?