Stochastic Adaptive Control (02421)

Lecture 2

Tobias K. S. Ritschel

Assistant Professor

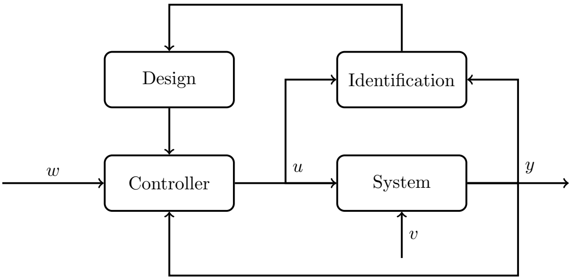
Section for Dynamical Systems, DTU Compute

DTU Compute

Department of Applied Mathematics and Computer Science

Lecture Plan

1. ARX models
2. ARX prediction + control
3. ARX estimation
4. ARX model validation + adaptive control
5. ARMAX control
6. ARMAX estimation + adaptive control
7. Systems and control theory
8. Stochastic systems + Kalman filtering
9. SS estimation (recursive) + control
10. SS control
11. SS estimation (batch)
12. SS estimation (recursive)
13. SS nonlinear control



Today's Agenda

- Minimum variance control for ARX models
- Least-squares (LS) estimation
- Maximum likelihood (ML) estimation
- Estimation for ARX models
- Recursive estimation

Control

Control of ARX processes

ARX process

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

General control law

$$R(q^{-1})u_t = Q(q^{-1})w_t - S(q^{-1})y_t \tag{2}$$

Criterion used to derive optimal control laws

$$\min_{u_t} J_t(y_{t+k}, u_t) \tag{3}$$

Minimum variance control

Minimum Variance Control

ARX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t \tag{4}$$

B is stable

Minimize the variance

$$J_t = \mathbb{E}[y_{t+k}^2] \tag{5}$$

$G_k(q^{-1})$ and $S_k(q^{-1})$ are the solution to the simple Diophantine equation

$$1 = A(q^{-1})G_k(q^{-1}) + q^{-k}S_k(q^{-1}) \tag{6}$$

Minimum Variance Control

m-step prediction and error

$$\hat{y}_{t+m} = B(q^{-1})G_m(q^{-1})u_{t+m-k} + S_m(q^{-1}y_t), \tag{7}$$

$$\tilde{y}_{t+m} = G_m(q^{-1})e_{t+m} \tag{8}$$

Cost function (we exploit that $\hat{y}_{t+m} \perp \tilde{y}_{t+m}$)

$$J_t = \mathbb{E}[y_{t+k}^2] = \mathbb{E}\left[\left(B(q^{-1})G_k(q^{-1})u_t + S_k(q^{-1})y_t\right)^2\right] \tag{9}$$

$$+ \mathbb{E}\left[\left(G_k(q^{-1})e_{t+k}\right)^2\right] \tag{10}$$

Optimal control law

$$B(q^{-1})G_k(q^{-1})u_t = -S_k(q^{-1})y_t \tag{11}$$

Minimum Variance Control

Closed-loop system

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t \quad (12)$$

$$= -q^{-k}B(q^{-1})\frac{S_k(q^{-1})}{B(q^{-1})G_k(q^{-1})}y_t + e_t \quad (13)$$

$$= -q^{-k}\frac{S_k(q^{-1})}{G_k(q^{-1})}y_t + e_t \quad (14)$$

Simplify

$$\overbrace{\left(G_k(q^{-1})A(q^{-1}) + q^{-k}S_k(q^{-1})\right)}^{=1 \text{ (Diophantine equation)}} y_t = G_k(q^{-1})e_t \quad (15)$$

Control law

$$u_t = \frac{-S_k(q^{-1})}{B(q^{-1})G_k(q^{-1})}y_t \quad (16)$$

$$= \frac{-S_k(q^{-1})}{B(q^{-1})G_k(q^{-1})}G_k(q^{-1})e_t \quad (17)$$

$$= \frac{-S_k(q^{-1})}{B(q^{-1})}e_t \quad (18)$$

Minimum Variance Control

Stationary closed-loop system

$$y_t = G_k(q^{-1})e_t, \qquad B(q^{-1})u_t = -S_k(q^{-1})e_t \tag{19}$$

The closed loop poles are determined by $B(q^{-1})$

The minimum variance controller has the following shortcomings

1. No possibility for setpoints
2. Large control effort
3. Undamped zeros (zeros outside of the unit circle)

$\mathbf{MV}_0$ controller

MV_0 controller

ARX process

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t \tag{20}$$

Cost function

$$J_t = \mathbb{E}[(y_{t+k} - w_t)^2] \tag{21}$$

Optimal control law

$$B(q^{-1})G_k(q^{-1})u_t = w_t - S_k(q^{-1})y_t \tag{22}$$

The Diophantine equation is the same as for the minimum variance controller

$$1 = A(q^{-1})G_k(q^{-1}) + q^{-k}S_k(q^{-1}) \tag{23}$$

MV$_0$ controller

Closed-loop system

$$y_t = q^{-k} w_t + G_k(q^{-1}) e_t, \tag{24}$$
$$B(q^{-1}) u_t = A(q^{-1}) w_t - S_k(q^{-1}) e_t \tag{25}$$

The poles are determined by $B(q^{-1})$

If $w_t = 0$, the MV$_0$ control becomes the minimum variance control

The MV_0 controller still has the following shortcomings.

1. Large control effort.
2. Undamped zeros.

Estimation

Estimation Methods

Observation equation

$$Y = G(\theta) + e, \quad y_t = g(t, \theta) + e_t \tag{26}$$

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

Residuals

$$\epsilon = Y - G(\hat{\theta}), \quad \epsilon_t = y_t - g(t, \hat{\theta}) \tag{27}$$

Linear case

$$G = \Phi\theta, \quad g(t, \theta) = \phi_t^T \theta \tag{28}$$

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

Least-squares estimation

Least Squares Method

Least squares (LS)

$$\min_{\theta} J_N(\theta) = \min_{\theta} \frac{1}{2}\sum_{t=1}^{N} \epsilon_t^2 = \min_{\theta} \frac{1}{2}\epsilon^T \epsilon \tag{29}$$

Solution

$$\left(\frac{\partial G(\theta)}{\partial \theta}\right)^T G(\theta) = \left(\frac{\partial G(\theta)}{\partial \theta}\right)^T Y \tag{30}$$

Linear case

$$\Phi^T \Phi \theta = \Phi^T Y, \quad \sum_{t=1}^{N} \phi_t \phi_t^T \theta = \sum_{t=1}^{N} \phi_t y_t \tag{31}$$

where Φ is

$$\Phi = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_N^T \end{bmatrix} \tag{32}$$

Least Squares Method

Parameter estimate

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y = \left(\sum_{t=1}^{N} \phi_t \phi_t^T \right)^{-1} \sum_{t=1}^{N} \phi_t y_t \tag{33}$$

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

Distribution of estimate

$$\hat{\theta} \sim \mathcal{F}(\theta, P_\theta), \tag{34}$$

$$P_\theta = \mathrm{Cov}(\hat{\theta}) = (\Phi^T \Phi)^{-1} \Phi^T P \Phi (\Phi^T \Phi)^{-1} \tag{35}$$

Uncorrelated noise ($\Sigma = I$)

$$P_\theta = \sigma^2 (\Phi^T \Phi)^{-1} \tag{36}$$

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

$$\mathrm{Cov}(\hat{\theta}) \approx \hat{\sigma}^2 \left(\frac{\partial^2 J_N}{\partial \theta^2} (\hat{\theta}) \right)^{-1}, \qquad \hat{\sigma}^2 \approx 2 \frac{J_N(\hat{\theta})}{N - n_\theta} \tag{37}$$

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 $\mathrm{Cov}(\hat{\theta}) = (\Phi^T\Phi)^{-1}\Phi^T P \Phi(\Phi^T\Phi)^{-1}$
- It does not assume a specific distribution

If $P = \sigma^2 I$

- Unbiased: $\mathbb{E}[\hat{\theta}] = \theta$ and $\mathrm{Cov}(\hat{\theta}) = \sigma^2(\Phi^T\Phi)^{-1}$
- 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

Linear Models

If we consider the parameters

$$\theta^T = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix} \tag{38}$$

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>two minutes</u> and
then discuss with the person next to you for <u>five minutes</u>.

Least squares – Example

System

$$y_t = \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + e_t \tag{39}$$

$$= \underbrace{\begin{bmatrix} u_t & u_{t-1} & u_{t-2} \end{bmatrix}}_{\phi_t^T} \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}}_{\theta} + e_t \tag{40}$$

Matrix ($N = 3$ measurements)

$$\Phi = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \phi_3^T \end{bmatrix} = \begin{bmatrix} u_1 & u_0 & u_{-1} \\ u_2 & u_1 & u_0 \\ u_3 & u_2 & u_1 \end{bmatrix} \tag{41}$$

Least squares – Example

Measurements

$$y_1 = 1, \qquad y_2 = 2, \qquad y_3 = 3 \tag{42}$$

Inputs

$$u_{-1} = 3, \qquad u_0 = 1, \qquad u_1 = 4, \qquad u_2 = -1, \qquad u_3 = 2 \tag{43}$$

Matrix

$$\Phi = \begin{bmatrix} 4 & 1 & 3 \\ -1 & 4 & 1 \\ 2 & -1 & 4 \end{bmatrix}, \qquad Y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{44}$$

Parameter estimate

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = (\Phi^T \Phi)^{-1} \Phi^T Y = \begin{bmatrix} -0.5660 \\ 0.0943 \\ 1.0566 \end{bmatrix} \tag{45}$$

Least squares – Exercise

Measurements

$$y_1 = 4, \qquad y_2 = -1, \qquad y_3 = 2 \tag{46}$$

Inputs

$$u_{-1} = 5, \qquad u_0 = 2, \qquad u_1 = -2, \qquad u_2 = -3, \qquad u_3 = 1 \tag{47}$$

Solve the exercise in 10 min.

Matrix

$$\Phi = \begin{bmatrix} -2 & 2 & 5 \\ -3 & -2 & 2 \\ 1 & -3 & -2 \end{bmatrix}, \qquad Y = \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix} \tag{48}$$

Parameter estimate

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = (\Phi^T \Phi)^{-1} \Phi^T Y = \begin{bmatrix} 2.9259 \\ -1.3704 \\ 2.5185 \end{bmatrix} \tag{49}$$

Maximum likelihood estimation

Maximum Likelihood Method

Likelihood

$$\mathcal{L}(\theta) = f(Y|\theta) \tag{50}$$

Maximum likelihood estimation problem (equivalent formulations)

$$\max_{\theta} \mathcal{L}(\theta), \qquad \max_{\theta} \ln \mathcal{L}(\theta) \tag{51}$$

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

$$f(Y|\theta) = \frac{1}{\sqrt{(2\pi)^N}\sqrt{\det P}} \exp\left(-\frac{1}{2}(Y - \Phi\theta)^T P^{-1}(Y - \Phi\theta)\right) \quad (52)$$

Log-likelihood function

$$\ln \mathcal{L}(Y;\theta) = -\frac{1}{2}\ln\det P - \frac{N}{2}\ln 2\pi - \frac{1}{2}(Y - \Phi\theta)^T P^{-1}(Y - \Phi\theta) \quad (53)$$

Optimization problem

$$\max_\theta \; \ln \mathcal{L}(\theta) = \min_\theta \; -\ln \mathcal{L}(\theta) \quad (54)$$

$$= \min_\theta \frac{1}{2}\ln\det P + \frac{1}{2}(Y - \Phi\theta)^T P^{-1}(Y - \Phi\theta) + c \quad (55)$$

c is constant and independent of θ and P

Maximum Likelihood Method

First-order optimality conditions

$$\frac{\partial \ln \mathcal{L}}{\partial \theta}(Y;\theta) = \frac{1}{2}(-2\Phi^T P^{-1} Y + 2\Phi^T P^{-1}\Phi\theta) = 0 \tag{56}$$

Optimal estimate

$$\hat{\theta} = (\Phi^T P^{-1}\Phi)^{-1}\Phi^T P^{-1} Y \tag{57}$$

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

$$\hat{\theta} = (\Phi^T \Sigma^{-1}\Phi)^{-1}\Phi^T \Sigma^{-1} Y \tag{58}$$

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

$$\hat{\theta} = \frac{\sigma^2}{\sigma^2}(\Phi^T\Phi)^{-1}\Phi^T Y = (\Phi^T\Phi)^{-1}\Phi^T Y \tag{59}$$

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

$$\frac{\partial \ln \mathcal{L}}{\partial \sigma^2}(Y;\theta) = \frac{N}{2\sigma^2} - \frac{1}{2\sigma^4}(Y - \Phi\theta)^T \Sigma^{-1}(Y - \Phi\theta) = 0 \tag{60}$$

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

ML estimate of the noise covariance

$$\hat{\sigma}^2 = \frac{(Y - \Phi\hat{\theta})^T \Sigma^{-1}(Y - \Phi\hat{\theta})}{N} \tag{61}$$

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

$$\begin{aligned}
\mathrm{Cov}(\epsilon, \hat{\theta}) &= \mathrm{Cov}(Y - \Phi\hat{\theta}, \hat{\theta}) && (62) \\
&= \mathrm{Cov}(\Phi\theta + e - \Phi\hat{\theta}, \hat{\theta}), \quad Y = \Phi\theta + e && (63) \\
&= \mathrm{Cov}(e, \hat{\theta}) - \Phi\,\mathrm{Cov}(\hat{\theta}, \hat{\theta}) && (64) \\
&= \mathrm{Cov}(e, e)L^T - \Phi L\,\mathrm{Cov}(e, e)L^T, \quad \hat{\theta} = LY = L\Phi\theta + Le && (65) \\
&= (I - \Phi L)PL^T && (66)
\end{aligned}$$

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

$$\mathrm{Cov}(\epsilon, \hat{\theta}) = (I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)P\Phi(\Phi^T\Phi)^{-1} \qquad (67)$$

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

$$\begin{aligned}
\mathrm{Cov}(\epsilon, \hat{\theta}) &= \sigma^2(I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)\Phi(\Phi^T\Phi)^{-1} && (68) \\
&= \sigma^2(\Phi(\Phi^T\Phi)^{-1} - \Phi(\Phi^T\Phi)^{-1}\Phi^T\Phi(\Phi^T\Phi)^{-1}) = 0 && (69)
\end{aligned}$$

ARX estimation

ARX estimation

ARX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t, \tag{70}$$

$$y_t = -\sum_{i=1}^{n_a} a_i y_{t-i} + \sum_{i=0}^{n_b} b_i u_{t-i-k} + e_t \tag{71}$$

$e_t \sim F(0, \sigma^2)$ and white

Rewrite

$$y_t = \sum_{i=1}^{n_\phi} \theta_i \phi_{t,i} + e_t = \phi_t^T \theta + e_t, \tag{72}$$

$$\phi_t^T = [-y_{t-1}, -y_{t-2}, ..., -y_{t-n_a}, u_{t-k}, ..., u_{t-k-n_b}], \tag{73}$$

$$\theta^T = [a_1, a_2, ..., a_{n_a}, b_0, b_1, ..., b_{n_b}] \tag{74}$$

Estimation - LS and ML

Least-squares method

$$Y_t = \Phi_t\theta + E_t, \quad E_t \sim \mathbb{F}(0, P) \tag{75}$$

$$\hat{\theta} = (\Phi_t^T\Phi_t)^{-1}\Phi_t^T Y_t \tag{76}$$

$$\Rightarrow \hat{\theta} \sim \mathbb{F}(\theta, (\Phi_t^T\Phi_t)^{-1}\Phi_t^T P\Phi_t(\Phi_t^T\Phi_t)^{-1}) \tag{77}$$

Maximum-likelihood method

$$Y_t = \Phi_t\theta + E_t, \quad E_t \sim N(0, P) \tag{78}$$

$$\hat{\theta} = (\Phi_t^T P^{-1}\Phi_t)^{-1}\Phi_t^T P^{-1} Y_t \tag{79}$$

$$\Rightarrow \hat{\theta} \sim \mathbb{F}(\theta, (\Phi_t^T P^{-1}\Phi_t)^{-1}\Phi_t^T P^{-1}\Phi_t(\Phi_t^T P^{-1}\Phi_t)^{-1}) \tag{80}$$

If $P = \sigma^2\Sigma$

$$\hat{\sigma}^2 = \frac{(Y - \Phi_t\hat{\theta})^T\Sigma^{-1}(Y - \Phi_t\hat{\theta})}{N} \tag{81}$$

Recursive parameter estimation

Recursive Estimation

The previously presented methods are in the form

$$\hat{\theta}_t = \text{func}(Y_t) \tag{82}$$

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

$$\hat{\theta}_t = \text{func}(y_t, \hat{\theta}_{t-1}) \tag{83}$$

- 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

$$\hat{\theta}_t = (\Phi^T \Phi)^{-1} \Phi^T Y_t \tag{84}$$

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

$$\hat{\theta}_t = \bar{\theta} + (\Phi^T \Phi)^{-1} \Phi^T \epsilon \tag{85}$$

Iterative formulation of the LS estimator

$$\hat{\theta}_t = \hat{\theta}_{t-1} + (\Phi_t^T \Phi_t)^{-1} \Phi_t^T \epsilon_t \tag{86}$$

RLS/RARX - Recursive Least Squares

ARX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t, \quad e_t \sim \mathcal{F}(0,\sigma^2) \tag{87}$$

$$y_t = \phi_t^T\theta + e_t, \quad e_t \perp e_s \quad s > t \tag{88}$$

$$\phi_t = [-y_{t-1}, \ldots, -y_{t-n_a}, u_{t-k}, \ldots, u_{t-n_b-k}]^T \tag{89}$$

$$\theta = [a_1, \ldots, a_{n_a}, b_0, \ldots, b_{n_b}]^T \tag{90}$$

Least squares estimator based on t measurements

$$\hat{\theta}_t = \left(\sum_{i=1}^{t} \phi_i\phi_i^T\right)^{-1} \sum_{i=1}^{t} \phi_i y_i, \tag{91}$$

$$P_t^{-1} = \sum_{i=1}^{t} \phi_i\phi_i^T, \quad \sum_{i=1}^{t} \phi_i\epsilon_i = 0 \tag{92}$$

RLS/RARX - Recursive Least Squares

Recursive formulation

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \sum_{i=1}^{t} \phi_i \epsilon_i \quad (93)$$

Rewrite the recursion

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \phi_t \epsilon_t \quad (94)$$

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \quad (95)$$

$$P_t^{-1} = P_{t-1}^{-1} + \phi_t \phi_t^T \quad (96)$$

$$\mathrm{Var}(\hat{\theta}_t | Y_t) = P_t \sigma^2 \approx \mathrm{Var}(\hat{\theta}_t) \quad (97)$$

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

$$\hat{\theta}_0 = 0, \quad P_0 = \beta I, \quad \beta \gg 0 \quad (98)$$

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)

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \tag{99}$$

$$s_t = 1 + \phi_t^T P_{t-1} \phi_t \tag{100}$$

$$K_t = \frac{P_{t-1}\phi_t}{s_t} \tag{101}$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t \epsilon_t \tag{102}$$

$$P_t = P_{t-1} - K_t s_t K_t^T \tag{103}$$

Questions?