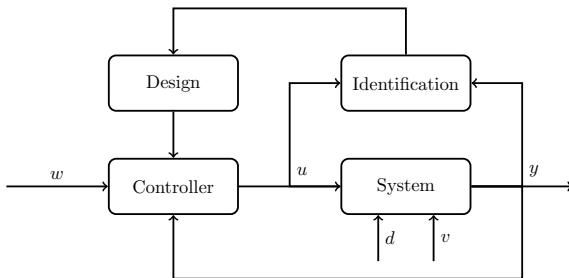


- ① ARX models
- ② ARX estimation
- ③ ARMAX estimation
- ④ ARMAX identification
- ⑤ ARMAX control
- ⑥ ARMAX adaptive control
- ⑦ Systems and control theory
- ⑧ Stochastic systems + SS estimation
- ⑨ SS estimation + control
- ⑩ SS control
- ⑪ To be decided ...
- ⑫ SS nonlinear estimation
- ⑬ SS nonlinear control



Today's Agenda

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

Estimation

Observation equation

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

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

Residuals

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

Linear case

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

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

Least-squares estimation

Least Squares Method

Least squares (LS)

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

Solution

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

Linear case

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

where Φ is

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

Least Squares Method

Parameter estimate

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

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

Distribution of estimate

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

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

Uncorrelated noise ($\Sigma = I$)

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

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

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

Residual property

$$\Phi^T \epsilon = 0, \quad \epsilon = Y - \Phi \hat{\theta} \quad (13)$$

Proof

$$\begin{aligned} \Phi^T \epsilon &= \Phi^T (Y - \Phi \hat{\theta}) = \Phi^T (Y - \Phi (\Phi^T \Phi)^{-1} \Phi^T Y) \\ &= \Phi^T Y - \Phi^T \Phi (\Phi^T \Phi)^{-1} \Phi^T Y \\ &= \Phi^T Y - \Phi^T Y = 0 \end{aligned} \quad (14)$$

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
- Residuals: $\Phi^T \epsilon = 0$

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

System

$$y_t = \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + e_t \quad (15)$$

$$= \underbrace{\begin{bmatrix} u_t & u_{t-1} & u_{t-2} \end{bmatrix}}_{\phi_t^T} \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}}_{\theta} + e_t \quad (16)$$

Matrix ($N = 3$ measurements)

$$\Phi = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \phi_3^T \end{bmatrix} = \begin{bmatrix} u_1 & u_0 & u_{-1} \\ u_2 & u_1 & u_0 \\ u_3 & u_2 & u_1 \end{bmatrix} \quad (17)$$

Least squares – Example

Measurements

$$y_1 = 1, \quad y_2 = 2, \quad y_3 = 3 \quad (18)$$

Inputs

$$u_{-1} = 3, \quad u_0 = 1, \quad u_1 = 4, \quad u_2 = -1, \quad u_3 = 2 \quad (19)$$

Matrix

$$\Phi = \begin{bmatrix} 4 & 1 & 3 \\ -1 & 4 & 1 \\ 2 & -1 & 4 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad (20)$$

Parameter estimate

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = (\Phi^T \Phi)^{-1} \Phi^T Y = \begin{bmatrix} -0.5660 \\ 0.0943 \\ 1.0566 \end{bmatrix} \quad (21)$$

Maximum likelihood estimation

Likelihood

$$\mathcal{L}(\theta) = f(Y|\theta) \quad (22)$$

Maximum likelihood estimation problem (equivalent formulations)

$$\max_{\theta} \mathcal{L}(\theta), \quad \max_{\theta} \ln \mathcal{L}(\theta) \quad (23)$$

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

$$f(Y|\theta) = \frac{1}{\sqrt{(2\pi)^N \det P}} \exp \left(-\frac{1}{2} (Y - \Phi\theta)^T P^{-1} (Y - \Phi\theta) \right) \quad (24)$$

Log-likelihood function

$$\ln \mathcal{L}(Y; \theta) = -\frac{1}{2} \ln \det P - \frac{N}{2} \ln 2\pi - \frac{1}{2} (Y - \Phi\theta)^T P^{-1} (Y - \Phi\theta) \quad (25)$$

Optimization problem

$$\max_{\theta} \ln \mathcal{L}(\theta) = \min_{\theta} -\ln \mathcal{L}(\theta) \quad (26)$$

$$= \min_{\theta} \frac{1}{2} \ln \det P + \frac{1}{2} (Y - \Phi\theta)^T P^{-1} (Y - \Phi\theta) + c \quad (27)$$

c is constant and independent of θ and P

First-order optimality conditions

$$\frac{\partial \ln \mathcal{L}}{\partial \theta}(Y; \theta) = \frac{1}{2}(-2\Phi^T P^{-1}Y + 2\Phi^T P^{-1}\Phi\theta) = 0 \quad (28)$$

Optimal estimate

$$\hat{\theta} = (\Phi^T P^{-1}\Phi)^{-1}\Phi^T P^{-1}Y \quad (29)$$

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

$$\hat{\theta} = (\Phi^T \Sigma^{-1}\Phi)^{-1}\Phi^T \Sigma^{-1}Y \quad (30)$$

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

$$\hat{\theta} = \frac{\sigma^2}{\sigma^2}(\Phi^T \Phi)^{-1}\Phi^T Y = (\Phi^T \Phi)^{-1}\Phi^T Y \quad (31)$$

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

First-order optimality conditions for σ^2

$$\frac{\partial \ln \mathcal{L}}{\partial \sigma^2}(Y; \theta) = \frac{N}{2\sigma^2} - \frac{1}{2\sigma^4}(Y - \Phi\theta)^T \Sigma^{-1}(Y - \Phi\theta) = 0 \quad (32)$$

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

ML estimate of the noise covariance

$$\hat{\sigma}^2 = \frac{(Y - \Phi\hat{\theta})^T \Sigma^{-1}(Y - \Phi\hat{\theta})}{N} \quad (33)$$

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

$$\text{Cov}(\epsilon, \hat{\theta}) = \text{Cov}(Y - \Phi\hat{\theta}, \hat{\theta}) \quad (34)$$

$$= \text{Cov}(\Phi\theta + e - \Phi\hat{\theta}, \hat{\theta}), \quad Y = \Phi\theta + e \quad (35)$$

$$= \text{Cov}(e, \hat{\theta}) - \Phi \text{Cov}(\hat{\theta}, \hat{\theta}) \quad (36)$$

$$= \text{Cov}(e, e)L^T - \Phi L \text{Cov}(e, e)L^T, \quad \hat{\theta} = LY = L\Phi\theta + Le \quad (37)$$

$$= (I - \Phi L)PL^T \quad (38)$$

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

$$\text{Cov}(\epsilon, \hat{\theta}) = (I - \Phi(\Phi^T \Phi)^{-1} \Phi^T)P\Phi(\Phi^T \Phi)^{-1} \quad (39)$$

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

$$\text{Cov}(\epsilon, \hat{\theta}) = \sigma^2(I - \Phi(\Phi^T \Phi)^{-1} \Phi^T)\Phi(\Phi^T \Phi)^{-1} \quad (40)$$

$$= \sigma^2(\Phi(\Phi^T \Phi)^{-1} - \Phi(\Phi^T \Phi)^{-1} \Phi^T \Phi(\Phi^T \Phi)^{-1}) = 0 \quad (41)$$

ARX estimation

ARX model

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t + d, \quad (42)$$

$$y_t = -\sum_{i=1}^{n_a} a_i y_{t-i} + \sum_{i=0}^{n_b} b_i u_{t-i-k} + e_t + d \quad (43)$$

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

Rewrite

$$y_t = \sum_{i=1}^{n_\phi} \theta_i \phi_{t,i} + e_t = \phi_t^T \theta + e_t, \quad (44)$$

$$\phi_t^T = [-y_{t-1}, -y_{t-2}, \dots, -y_{t-n_a}, u_{t-k}, \dots, u_{t-k-n_b}, 1], \quad (45)$$

$$\theta^T = [a_1, a_2, \dots, a_{n_a}, b_0, b_1, \dots, b_{n_b}, d] \quad (46)$$

Least-squares method

$$Y_t = \Phi_t \theta + E_t, \quad E_t \sim \mathbb{F}(0, P_t) \quad (47)$$

$$\hat{\theta} = (\Phi_t^T \Phi_t)^{-1} \Phi_t^T Y_t \quad (48)$$

Maximum-likelihood method

$$Y_t = \Phi_t \theta + E_t, \quad E_t \sim N(0, P_t) \quad (49)$$

$$\hat{\theta} = (\Phi_t^T P^{-1} \Phi_t)^{-1} \Phi_t^T P^{-1} Y_t \quad (50)$$

Approximate distribution of least-squares estimate

$$\hat{\theta} \sim \mathbb{F}(\theta, (\Phi_t^T \Phi_t)^{-1} \Phi_t^T P \Phi_t (\Phi_t^T \Phi_t)^{-1}) \quad (51)$$

Approximate distribution of maximum likelihood estimates

$$\hat{\theta} \sim \mathbb{F}(\theta, (\Phi_t^T P_t^{-1} \Phi_t)^{-1} \Phi_t^T P_t^{-1} \Phi_t (\Phi_t^T P_t^{-1} \Phi_t)^{-1}) \quad (52)$$

If $P_t = \sigma^2 \Sigma_t$

$$\hat{\sigma}^2 = \frac{(Y_t - \Phi_t \hat{\theta})^T \Sigma_t^{-1} (Y_t - \Phi_t \hat{\theta})}{N} \quad (53)$$

Recursive parameter estimation

The previously presented methods are in the form

$$\hat{\theta}_t = \text{func}(Y_t) \quad (54)$$

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

$$\hat{\theta}_t = \text{func}(y_t, \hat{\theta}_{t-1}) \quad (55)$$

- 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

$$\hat{\theta}_t = (\Phi_t^T \Phi_t)^{-1} \Phi_t^T Y_t, \quad \epsilon_{t|t-1} = y_t - \phi_t^T \hat{\theta}_{t-1} \quad (56)$$

Residuals

$$R_{t|t-1} = Y_t - \Phi_t \hat{\theta}_{t-1} = \begin{bmatrix} R_{t-1|t-1} \\ \epsilon_{t|t-1} \end{bmatrix}, \quad R_{t-1|t-1} = Y_{t-1} - \Phi_{t-1} \hat{\theta}_{t-1} \quad (57)$$

Write out least squares estimator

$$\begin{aligned} \hat{\theta}_t &= (\Phi_t^T \Phi_t)^{-1} \Phi_t^T (\Phi_t \hat{\theta}_{t-1} + R_{t|t-1}) = \hat{\theta}_{t-1} + (\Phi_t^T \Phi_t)^{-1} \Phi_t^T R_{t|t-1} \\ &= \hat{\theta}_{t-1} + (\Phi_t^T \Phi_t)^{-1} \begin{bmatrix} \Phi_{t-1}^T & \phi_t^T \end{bmatrix} \begin{bmatrix} R_{t-1|t-1} \\ \epsilon_{t|t-1} \end{bmatrix} \\ &= \hat{\theta}_{t-1} + (\Phi_t^T \Phi_t)^{-1} \phi_t \epsilon_{t|t-1} = \hat{\theta}_{t-1} + P_t \phi_t \epsilon_{t|t-1} \end{aligned} \quad (58)$$

Write out inverse of variance

$$\begin{aligned} P_t^{-1} &= \Phi_t^T \Phi_t = \begin{bmatrix} \Phi_{t-1}^T & \phi_t^T \end{bmatrix} \begin{bmatrix} \Phi_{t-1} \\ \phi_t \end{bmatrix} = \Phi_{t-1}^T \Phi_{t-1} + \phi_t \phi_t^T \\ &= P_{t-1}^{-1} + \phi_t \phi_t^T \end{aligned} \quad (59)$$

Recursion

$$\hat{\theta}_t = \hat{\theta}_{t-1} + P_t \phi_t \epsilon_t \quad (60)$$

$$\epsilon_{t|t-1} = y_t - \phi_t^T \hat{\theta}_{t-1} \quad (61)$$

$$P_t^{-1} = P_{t-1}^{-1} + \phi_t \phi_t^T \quad (62)$$

$$\text{Var}(\hat{\theta}_t | Y_t) = P_t \sigma^2 \approx \text{Var}(\hat{\theta}_t) \quad (63)$$

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

$$\hat{\theta}_0 = 0, \quad P_0 = \beta I, \quad \beta \gg 0 \quad (64)$$

The recursion can also be computed using alternative formulations

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

$$\epsilon_t = y_t - \phi_t^T \hat{\theta}_{t-1} \quad (65)$$

$$s_t = 1 + \phi_t^T P_{t-1} \phi_t \quad (66)$$

$$K_t = \frac{P_{t-1} \phi_t}{s_t} \quad (67)$$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + K_t \epsilon_t \quad (68)$$

$$P_t = P_{t-1} - K_t s_t K_t^T \quad (69)$$

Questions?