Stochastic Adaptive Control (02421)

Lecture 9

Tobias K. S. Ritschel

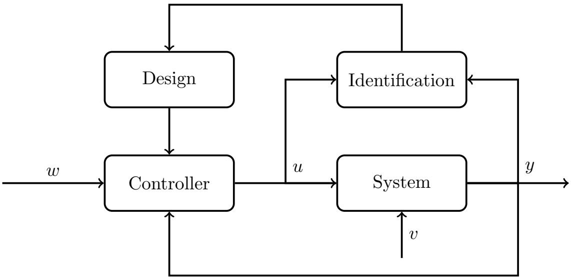
Section for Dynamical Systems, DTU Compute

DTU Compute

Department of Applied Mathematics and Computer Science

Lecture Plan

1. Systems theory
2. Stochastics
3. State estimation - Kalman filter 1
4. State estimation - Kalman filter 2
5. Optimal control 1 - internal models
6. External models
7. Prediction + optimal control 1 - external models
8. Optimal control 2 - external models
9. **System identification 1**
10. System identification 2
11. System identification 3 + model validation
12. Adaptive control 1
13. Adaptive control 2



Today's Agenda

- Follow-up from last lecture
- Estimation - ARX
- Estimation - ARMAX
- Estimation - L-Structure
- Estimation - Matlab commands

Follow-up from last time, Controller Implementation

If we have designed an external controller in the form:

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

we can implement it in a loop by using a state-space approach:

```
for t=1:N
    y_t = measure()

    [Aw,Bw,Cw,Dw]=tf2ss(Q,R)
    [Ay,By,Cy,Dy]=tf2ss(S,R)
    [Ad,Bd,Cd,Dd]=tf2ss(1,R)

    Xw = Aw*Xw + Bw*w_t
    Xy = Ay*Xy + By*y_t
    Xd = Ad*Xd + Bd*d
    u_t = CwXw + Dw*w(t) + Cy*Xy + Dy*y_t + Cd*Xd + Dd*d

    apply(u_t)
end
```

Follow-up from last time, reminder of statistic parameters

If we consider $X_{t,j}$ to be a sample at time t during the jth simulation, each simulation being N steps long. Then, the stationary mean and variance are estimated from the M simulations by

$$\hat{E}\{X\} = \frac{1}{M}\sum_{j=1}^{M} X_{t,j} \tag{2}$$

$$\hat{V}\{X\} = \frac{1}{M-1}\sum_{j=1}^{M}(X_{t,j} - \hat{E}\{X\})^2 \tag{3}$$

If the system is ergodic, we can obtain the same information from only a single simulation j, but using multiple time steps:

$$\hat{E}\{X\} = \frac{1}{N}\sum_{t=1}^{N} X_{t,j} \tag{4}$$

$$\hat{V}\{X\} = \frac{1}{N-1}\sum_{t=1}^{N}(X_{t,j} - \hat{E}\{X\})^2 \tag{5}$$

Follow-up from last time

Questions?

Estimation Methods

In methods for parameter estimation we assume we have a vector of observation Y, and a relation $G(\theta)$ to the parameters θ

$$Y = G(\theta) + e, \quad y_t = g(t, \theta) + e_t \tag{6}$$

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

we will in our estimations consider the residuals of our estimate:

$$\epsilon = Y - G(\hat{\theta}), \quad \epsilon_t = y_t - g(t, \hat{\theta}) \tag{7}$$

In the linear case

$$G = \Phi\theta, \quad g(t, \theta) = \phi_t^T \theta \tag{8}$$

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

Least Squares Method

The first method is the least squares method (LS)

$$\min_{\theta} J_N(\theta) = \min_{\theta} \frac{1}{2}\sum_{t=1}^{N} \epsilon_t^2 = \min_{\theta} \frac{1}{2}\epsilon^T\epsilon \tag{9}$$

The general solution is then given by

$$\left(\frac{\partial G(\theta)}{\partial \theta}\right)^T G(\theta) = \left(\frac{\partial G(\theta)}{\partial \theta}\right)^T Y \tag{10}$$

where we can see the results relies on the structure of $G(\theta)$.
If we consider the linear case, we get:

$$\Phi^T\Phi\theta = \Phi^T Y, \quad \sum_{t=1}^{N} \phi_t\phi_t^T\theta = \sum_{t=1}^{N} \phi_t y_t \tag{11}$$

where Φ is

$$\Phi = \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_N^T \end{bmatrix} \tag{12}$$

Least Squares Method

From this we can get our estimate of the parameters:

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y = \left(\sum_{t=1}^{N} \phi_t \phi_t^T \right)^{-1} \sum_{t=1}^{N} \phi_t y_t \tag{13}$$

if $\Phi^T \Phi$ has full rank.
The estimate has the statistical property $\hat{\theta} \sim \mathcal{F}(\theta, P_\theta)$:

$$P_\theta = \text{Var}\left[\hat{\theta}\right] = (\Phi^T \Phi)^{-1} \Phi^T P \Phi (\Phi^T \Phi)^{-1} \tag{14}$$

If the noise is uncorrelated with equal variance, then $\Sigma = I$:

$$P_\theta = \sigma^2 (\Phi^T \Phi)^{-1} \tag{15}$$

If the noise variance is unknown, but assumed to follow the above case, then it can be estimated by

$$\text{Var}\left[\hat{\theta}\right] \approx \hat{\sigma}^2 \left[\frac{\partial^2}{\partial \theta^2} J_N(\hat{\theta}) \right]^{-1}, \quad \hat{\sigma}^2 \approx \frac{2 J_N(\hat{\theta})}{N - n_\phi} \tag{16}$$

Least Squares Method

Main Properties

Linear least squares estimates have the properties:

- It is a linear function of the observations, Y.
- It is unbiased: $\mathsf{E}\left[\hat{\theta}\right] = \theta$ and $\mathsf{Var}\left[\hat{\theta}\right] = (\Phi^T\Phi)^{-1}\Phi^T P\Phi(\Phi^T\Phi)^{-1}$
- holds no assumption on distribution

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

- unbiased: $\mathsf{E}\left[\hat{\theta}\right] = \theta$ and $\mathsf{Var}\left[\hat{\theta}\right] = \sigma^2(\Phi^T\Phi)^{-1}$
- Independence: $\epsilon \perp \hat{\theta}$
- $\hat{\theta}$ is BLUE (Best Linear Unbiased Estimator), which means that it has the smallest variance among all estimators which are linear functions of the observations.

Estimation

A linear parameterized model could be written as

$$y_t = a + bx_t + cx_{t-1}^2 + e_t = \begin{bmatrix} 1 & x_t & x_{t-1}^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \epsilon_t = \Phi(x_t)\theta + e_t \quad (17)$$

Being Unbiased or central estimator:

$$E\{\hat{\theta} - \theta\} = 0 \quad (18)$$

Being a minimal variance estimator

$$Var(\hat{\theta} - \theta|\theta) \leq Var(\bar{\theta} - \theta|\theta) \quad (19)$$

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

Linear Models?

If we consider the parameters

$$\theta^T = [\theta_1, \theta_2, \theta_3] \tag{20}$$

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>.

Maximum Likelihood Method

Let us now consider the Maximum Likelihood method (ML), where we consider the maximization of the likelihood function $\mathcal{L}$:

$$\mathcal{L}(\theta) = f(Y|\theta) \tag{21}$$

The minimum can be found in two ways:

$$\max_{\theta} \mathcal{L}(\theta), \quad \max_{\theta} \ln(\mathcal{L}(\theta)) \tag{22}$$

The likelihood function, $\mathcal{L}$, is the joint probability distribution function for all observations for given values of θ.

Estimation using ML, unlike LS, requires an assumption on the stochastic distribution of the residuals and the observations.

Maximum Likelihood Method

In the rest of our discussion, we will assume the system is linear, $Y = \Phi\theta + e$ as before, as well as normal-distributed.

Thus, we get that the joint-probability of N observations are

$$f(Y|\theta) = \frac{1}{\sqrt{Det(P)}\sqrt{(2\pi)^N}} \exp\left(-\frac{1}{2}(Y - \Phi\theta)^T P^{-1}(Y - \Phi\theta)\right) \tag{23}$$

If we apply the log-approach to the likelihood:

$$\ln \mathcal{L}(Y;\theta) = -\frac{1}{2}\ln\Big(Det(P)\Big) - \frac{N}{2}\ln\Big(2\pi\Big) - \frac{1}{2}(Y - \Phi\theta)^T P^{-1}(Y - \Phi\theta) \tag{24}$$

Then, we obtain the following optimization problem

$$\max_\theta \ln(\mathcal{L}(\theta)) = \min_\theta - \ln(\mathcal{L}(\theta)) \tag{25}$$

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

where c is a constant independent of θ and P.

Maximum Likelihood Method

The first-order optimality conditions for the estimate $\hat{\theta}$ is then given by

$$\frac{\partial \ln \mathcal{L}}{\partial \theta}(Y;\theta) = \frac{1}{2}(-2\Phi^T P^{-1} Y + 2\Phi^T P^{-1}\Phi\theta) = 0. \tag{26}$$

Solving for θ, optimality is obtained for

$$\hat{\theta} = (\Phi^T P^{-1}\Phi)^{-1}\Phi^T P^{-1} Y \tag{27}$$

Notice that if we have iid for the noise $(P = \sigma^2 I)$ it becomes the LS estimator:

$$\hat{\theta} = \frac{\sigma^2}{\sigma^2}(\Phi^T\Phi)^{-1}\Phi^T Y = (\Phi^T\Phi)^{-1}\Phi^T Y \tag{28}$$

Furthermore, we can see, that only the structure Σ of the variance $P = \sigma^2\Sigma$ is important, not the size σ^2

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

Maximum Likelihood Method

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

We can therefore estimate σ^2. The first-order optimality conditions for σ^2 are given by

$$\frac{\partial \ln \mathcal{L}}{\partial \sigma^2}(Y;\theta) = \frac{N}{2\sigma^2} - \frac{1}{2\sigma^4}(Y - \Phi\theta)^T \Sigma^{-1}(Y - \Phi\theta) = 0. \quad (30)$$

given $Det(P) = (\sigma^2)^N Det(\Sigma)$.

Solving for σ^2, optimality is obtained for

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

Maximum Likelihood Method

Main Properties

For the assumption of Normality, the Maximum Likelihood estimates have the properties:

- 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 $P = \sigma^2 I$

- estimate is LS
- unbiased: $\hat{\theta} \sim N\{\theta, \sigma^2(\Phi^T\Phi)^{-1}\}$
- Independence: $\epsilon \perp \hat{\theta}$
- $\hat{\theta}$ is BLUE (Best Linear Unbiased Estimator), which means that it has the smallest variance among all estimators which are linear functions of the observations.

Residual-Estimator Independence

We have now seen that both the LS and ML estimators, achieves Residual-Estimator Independence when $P = \sigma^2 I$.

Let us check this through the covariance:

$$
\begin{aligned}
cov(\epsilon, \hat{\theta}) &= cov(Y - \Phi\hat{\theta}, \hat{\theta}) && (32)\\
&= cov(\Phi\theta + e - \Phi\hat{\theta}, \hat{\theta}), \quad Y = \Phi\theta + e && (33)\\
&= cov(e, \hat{\theta}) - \Phi cov(\hat{\theta}, \hat{\theta}) && (34)\\
&= cov(e, e)L^T - \Phi L cov(e, e)L^T, \quad \hat{\theta} = LY = L\Phi\theta + Le && (35)\\
&= (I - \Phi L)PL^T && (36)
\end{aligned}
$$

If we apply the LS estimator $L = (\Phi^T\Phi)^{-1}\Phi^T$, we get

$$cov(\epsilon, \hat{\theta}) = (I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)P\Phi(\Phi^T\Phi)^{-1} \tag{37}$$

and we can then see that if P is a multiple of the identity matrix,

$$
\begin{aligned}
cov(\epsilon, \hat{\theta}) &= P(I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)\Phi(\Phi^T\Phi)^{-1} && (38)\\
&= P(\Phi(\Phi^T\Phi)^{-1} - \Phi(\Phi^T\Phi)^{-1}\Phi^T\Phi(\Phi^T\Phi)^{-1}) = 0 && (39)
\end{aligned}
$$

Estimation using different Model Framework

We will now consider how to apply the estimation methods for 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: $A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t$
- L: $A(q^{-1})y_t = q^{-k}\frac{B(q^{-1})}{F(q^{-1})}u_t + \frac{C(q^{-1})}{D(q^{-1})}e_t$

ARX estimation

We will now consider estimation of an ARX-model:

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t \tag{40}$$

$$y_t = -\sum_{i=1}^{n_a} a_i y_{t-i} + \sum_{i=0}^{n_b} b_i u_{t-i-k} + e_t, \tag{41}$$

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

Given the ARX linear form, we can rearrange it:

$$y_t = \sum_{i=1}^{n_\phi} \theta_i \phi_{t,i} + e_t = \phi_t^T \theta + e_t \tag{42}$$

$$\phi_t^T = [-y_{t-1}, -y_{t-2}, ..., -y_{t-n_a}, u_{t-k}, ..., u_{t-k-n_b}] \tag{43}$$

$$\theta^T = [a_1, a_2, ..., a_{n_a}, b_0, b_1, ..., b_{n_b}] \tag{44}$$

Estimation - LS and ML

Least-squares method:

$$Y_t = \Phi_t\theta + E_t, \quad E_t \in \mathbb{F}(0, P) \tag{45}$$

$$\hat{\theta} = (\Phi_t^T\Phi_t)^{-1}\Phi_t^T Y_t \tag{46}$$

$$\Rightarrow \hat{\theta} \in \mathbb{F}(\theta, (\Phi_t^T\Phi_t)^{-1}\Phi_t^T P\Phi_t(\Phi_t^T\Phi_t)^{-1}) \tag{47}$$

Maximum-likelihood method:

$$Y_t = \Phi_t\theta + E_t, \quad E_t \in N(0, P) \tag{48}$$

$$\hat{\theta} = (\Phi_t^T P^{-1}\Phi_t)^{-1}\Phi_t^T P^{-1} Y_t \tag{49}$$

$$\Rightarrow \hat{\theta} \in \mathbb{F}(\theta, (\Phi_t^T P^{-1}\Phi_t)^{-1}\Phi_t^T P^{-1}\Phi_t(\Phi_t^T P^{-1}\Phi_t)^{-1}) \tag{50}$$

$$\text{if: } P = \Sigma\sigma^2 \tag{51}$$

$$\Rightarrow \hat{\sigma}^2 = \frac{(Y - \Phi_t\hat{\theta})^T\Sigma^{-1}(Y - \Phi_t\hat{\theta})}{N} \tag{52}$$

ARX Example

We will consider the model

$$y_t = \frac{-1}{4} y_{t-1} + \frac{1}{2} y_{t-2} + u_{t-1} + e_t, \tag{53}$$

where e_t follows our usual assumptions with unit variance. We will let u_t be a integer random input sequence and thereby excite the system. Let us say we have 10 samples, and the system variance is $\sigma^2 = 0.1$:

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

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

Computing Φ and applying $\theta = (\Phi_t^T \Phi_t)^{-1} \Phi_t^T Y_t$

$$\theta^T = [0.2556, -0.5058, 0.9959], \quad P_\theta = \begin{bmatrix} 0.1287 & -0.0403 & 0.1010 \\ -0.0403 & 0.0599 & 0.0100 \\ 0.1010 & 0.0100 & 0.1203 \end{bmatrix}$$

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

Estimation - OE methods

In output-error (OE) estimation, we only consider the output-error of our model, or the input/output relation:

$$y_t = q^{-k}\frac{B(q^{-1})}{A(q^{-1})}u_t + e_t \tag{54}$$

where the estimation is based on minimization of the error:

$$J = \frac{1}{2}\sum_{t=1}^{N}\epsilon_t^2 = \frac{1}{2}\sum_{t=1}^{N}\left(y_t - q^{-k}\frac{B(q^{-1})}{A(q^{-1})}u_t\right)^2 \tag{55}$$

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

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)

The LS estimator has the benefit that the criteria is a square of the parameters. So we can estimate the input-output model using LS. But if we don't have an ARX model, the noise is non-white and LS is inconsistent. We can use the IV-method as alternative to the LS-method:

$$\hat{\theta} = \left(\sum_{t=1}^{N} \psi_t \phi_t \right)^{-1} \sum_{t=1}^{N} \psi_t^T y_t = \left(\Psi \Phi \right)^{-1} \Psi^T Y \tag{56}$$

where ψ_t is chosen such that:

$$E\{\psi_t e_t\} = 0 \tag{57}$$

$$E\{\psi_t \phi_t^T\} \text{ is invertible} \tag{58}$$

In practise $\psi_t \simeq \phi_t$ could, for instance, be chosen as

$$\phi_t = [-y_{t-1}, -y_{t-2}, ..., -y_{t-n_a}, u_t, u_{t-k}, ..., u_{t-k-n_b}] \tag{59}$$

$$\psi_t = [-\bar{y}_{t-1}, -\bar{y}_{t-2}, ..., -\bar{y}_{t-n_a}, u_t, u_{t-k}, ..., u_{t-k-n_b}] \tag{60}$$

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

Estimation - Estimation by Iteration

The LS-estimator were given as

$$\hat{\theta}_t = (\Phi^T \Phi)^{-1} \Phi^T Y_t \tag{61}$$

If we consider the output as $Y_t = \Phi\bar{\theta} + \epsilon$, for some previously estimator $\bar{\theta}$. Then we have

$$\hat{\theta}_t = \bar{\theta} + (\Phi^T \Phi)^{-1} \Phi^T \epsilon \tag{62}$$

This gives us an iterative formulation of the LS-estimator:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + (\Phi_t^T \Phi_t)^{-1} \Phi_t^T \epsilon_t \tag{63}$$

Estimation - Bayes estimation

If we again consider the ARX model:

$$A(q^{-1})y_t = B(q^{-1})u_t + e_t, \quad e_t \perp e_s, \quad e_t \in N(0, \sigma_t^2) \tag{64}$$

which we can write for time t and t+1 as

$$y_t = \phi_t^T \theta + e_t \tag{65}$$

$$y_{t+1} = \phi_{t+1}^T \theta + e_{t+1} \tag{66}$$

In Bayesian estimation, we consider the conditional estimator:

$$\theta | Y_t \in N(\hat{\theta}_t, P_t) \tag{67}$$

similarly to the estimation in the Kalman filter.

Estimation - Bayes estimation

Our estimator can then be formulated as

$$\hat{\theta}_{t+1} = \hat{\theta}_t + K_{t+1}(y_{t+1} - \phi_{t+1}^T \hat{\theta}_t) \tag{68}$$

$$K_{t+1} = \frac{P_{t+1}\phi_{t+1}}{\sigma_{t+1}^2} = P_t \phi_{t+1}(\phi_{t+1}^T P_t \phi_{t+1} + \sigma_{t+1}^2)^{-1} \tag{69}$$

$$P_{t+1}^{-1} = P_t^{-1} + \frac{\phi_{t+1}\phi_{t+1}^T}{\sigma_{t+1}^2}, \quad P_{t+1} = (I - K_{t+1}\phi_{t+1}^T)P_t \tag{70}$$

where our estimation error is given by:

$$\tilde{\theta}|Y_t \in N(0, P_t) \tag{71}$$

Due to the asymptotic behaviour of the estimator, we can approximately denote the error as:

$$\tilde{\theta}_t \,\tilde{\in}\, N(0, P_t) \tag{72}$$

ARMAX estimation

If we now consider the ARMAX models

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t \tag{73}$$

We can write the ARMAX model with the same structure as the ARX model:

$$y_t = \phi_t^T\theta + e_t \tag{74}$$

$$\phi_t = (-y_{t-1}, -y_{t-2}, ..., -y_{t-n_a}, u_{t-k}, u_{t-k-1}, ..., u_{t-k-n_b}, e_{t-1}, e_{t-2}, ..., e_{t-n_c})^T \tag{75}$$

Consider N consecutive measurements,

$$Y_N = \Phi_N\theta + E_N \tag{76}$$

$$Y_N = (y_1, ..., y_N)^T \quad E_N = (e_1, ..., e_N)^T \tag{77}$$

where Φ_N includes estimates of e.

ARMAX estimation - Extended Least-Square

With the previous form of the ARMAX model, it would be sensible to apply the LS method, but we will need to estimate the noise by the residuals ϵ:

$$e_t \approx \epsilon_t = y_t - \phi_t^T \hat{\theta} \tag{78}$$

where ϕ_t includes previous noise estimates. We can rewrite it as transfer functions:

$$\epsilon_t = \frac{\hat{A}(q^{-1})}{\hat{C}(q^{-1})} y_t - \frac{\hat{B}(q^{-1})}{\hat{C}(q^{-1})} u_t \tag{79}$$

Our estimator and error variance are then given by

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \left(\sum_{i=1}^{N} \phi_i \phi_i^T \right)^{-1} \sum_{i=1}^{N} \phi_i \epsilon_i \tag{80}$$

$$P_N = \frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2 \times \left(\sum_{i=1}^{N} \phi_i \phi_i^T \right)^{-1} \tag{81}$$

ARMAX estimation - Maximum Likelihood

We can do something similar when using the ML method for the ARMAX model,

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + C(q^{-1})e_t \tag{82}$$

where the residuals are again given as:

$$\epsilon_t = \frac{\hat{A}(q^{-1})}{\hat{C}(q^{-1})}y_t - \frac{\hat{B}(q^{-1})}{\hat{C}(q^{-1})}u_t = y_t - \phi_t^T\hat{\theta} \tag{83}$$

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

ARMAX estimation - Maximum Likelihood

Using the residuals, our estimator can then be designed using the approximate Newton method as

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \left(\sum_{i=1}^{N} \psi_i \psi_i^T \right)^{-1} \sum_{i=1}^{N} \psi_i \epsilon_i \tag{84}$$

where ψ_t is a filtered version of ϕ_t, given by

$$\psi_t = \frac{1}{\hat{C}(q^{-1})} \phi_t \tag{85}$$

where ψ_t can be computed through a state-space realization, or using past values:

$$\psi_t = \phi_t - \sum_{i=1}^{n_c} \hat{c}_i \psi_{t-i} \tag{86}$$

L-structure estimation - ELS method

If now we consider the general L-structure:

$$A(q^{-1})y_t = q^{-k}\frac{B(q^{-1})}{F(q^{-1})}u_t + \frac{C(q^{-1})}{D(q^{-1})}e_t + d \tag{87}$$

we can rewrite the structure in terms of filtered variables:

$$y_t = -\left(A(q^{-1}) - 1\right)y_t + B(q^{-1})u_{t-k} + \left(C(q^{-1}) - 1\right)e_t \tag{88}$$

$$-\left(F(q^{-1}) - 1\right)y_t^u - \left(D(q^{-1}) - 1\right)y_t^e + d + e_t \tag{89}$$

$$y_t^u = q^{-k}\frac{B(q^{-1})}{F(q^{-1})}u_t, \quad y_t^e = \frac{C(q^{-1})}{D(q^{-1})}e_t = A(q^{-1})y_t - y_t^u - d \tag{90}$$

We can now write the model as a linear model of the parameters:

$$y_t = \phi_t^T\theta + e_t \tag{91}$$

$$\phi_t = (-y_{t-1}, ...u_{t-k}, ..., -y_{t-1}^u, ..., e_{t-1}, ..., -y_{t-1}^e, ..., 1)^T \tag{92}$$

$$\theta = (a_1, ..., b_0, ..., f_1, ..., c_1, ..., d_1, ..., d)^T \tag{93}$$

We can now apply the ELS-method.

L-structure estimation - prediction error estimate (PEM-method)

If we consider the general L-structure:

$$A(q^{-1})y_t = q^{-k}\frac{B(q^{-1})}{F(q^{-1})}u_t + \frac{C(q^{-1})}{D(q^{-1})}e_t + d \tag{94}$$

If we also consider the first-step prediction:

$$y_{t+1} = \hat{y}_{t+1} + \tilde{y}_{t+1} = \hat{y}_{t+1} + e_{t+1} \tag{95}$$

$$\tilde{y}_{t+1} = Ge_{t+1}, \quad G(0) = 1, order(G) = 0 \tag{96}$$

we can rewrite the structure:

$$CF\hat{y}_t = F(C - AD)y_t + BDu_t + FDd \tag{97}$$

The PEM method is a generalised ML method, using the above structure to find the ψ-vectors (the error gradient):

$$\psi_t = \nabla_\theta \hat{y}_t = -\nabla \epsilon_t \tag{98}$$

L-structure estimation - prediction error estimate (PEM method)

The estimator and uncertainty is then given by

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \left(\sum_{i=1}^{N} \psi_i \psi_i^T \right)^{-1} \times \sum_{i=1}^{N} \psi_i \epsilon_i \tag{99}$$

$$P_N = \frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2 \times \left(\sum_{i=1}^{N} \psi_i \psi_i^T \right)^{-1} \tag{100}$$

and ψ is given by

$$\psi_t = (-\check{y}_{t-1}, ...\check{u}_t, ..., -\check{y}^u_{y-1}, ..., \check{e}_{t-1}, ..., -\check{y}^e_{t-1}, ..., \delta)^T \tag{101}$$

$$\check{y}_t = \frac{D}{C} y_t, \quad \check{u}_t = \frac{D}{CF} u_t, \quad \check{y}^u_t = -\frac{D}{CF} y^u_t \tag{102}$$

$$\check{e}_t = \frac{1}{C} \epsilon_t, \quad \check{y}^e_t = -\frac{1}{C} y^e_t, \quad \delta = \frac{D}{C} 1 \tag{103}$$

where y^u_t, y^e_t are given by (90).

Commands - Matlab

Commands that are good to know:

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

Questions

Questions?

Examples

Today's Matlab example topics:

- PRBS signals
- data generation
- estimation examples