Stochastic Adaptive Control (02421)

Lecture 1

Tobias K. S. Ritschel

Assistant Professor

Section for Dynamical Systems, DTU Compute

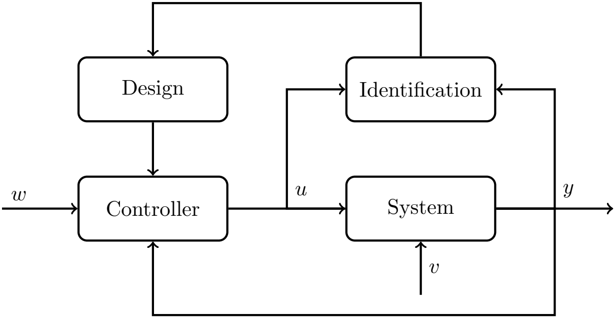
Course Content

Course details

- Time: Tuesday 08:00 - 12:00 (lectures and exercises)
- 5 ECTS points
- Evaluation: 4 individual reports
- Software: MATLAB (free choice)

Course plan

- Stochastic processes and systems (state space and transfer function models)
- Filter and control design
- System identification
- Adaptive control



Teachers

Tobias K. S. Ritschel
Assistant Professor
Course responsible
Contact: tobk@dtu.dk
Office: 303B-052

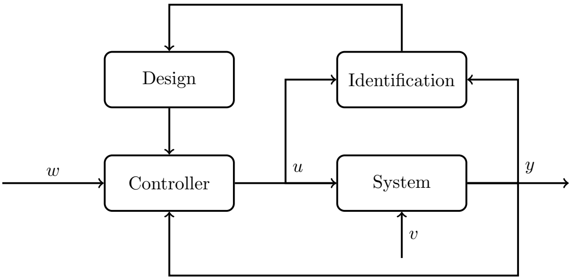
Henrik Madsen
Professor
Course co-responsible
Contact: hmad@dtu.dk
Office: 303B-004

You

- All of you are **MSc** students.
- A few guest students (abroad).
- Most of you are from **electrical** engineering (incl. autonomous systems) or **mathematical** engineering.
- A few from **chemical engineering**.

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



Structure of lectures

- Lectures and exercise sessions will be integrated
- Agenda + practical information
- Lecture content

Workload

Advice for this year

- If you're comfortable with the exercises, the mandatory assignments should be manageable. But consider working together with your fellow students, even though the report is individual.
- It's normal to feel stuck in this course. Therefore, ask questions!

New this year

- Revised structure of lectures
- More focused course content
- Smaller assignments (4 instead of 2)
- Fewer exercises

Matlab toolboxes

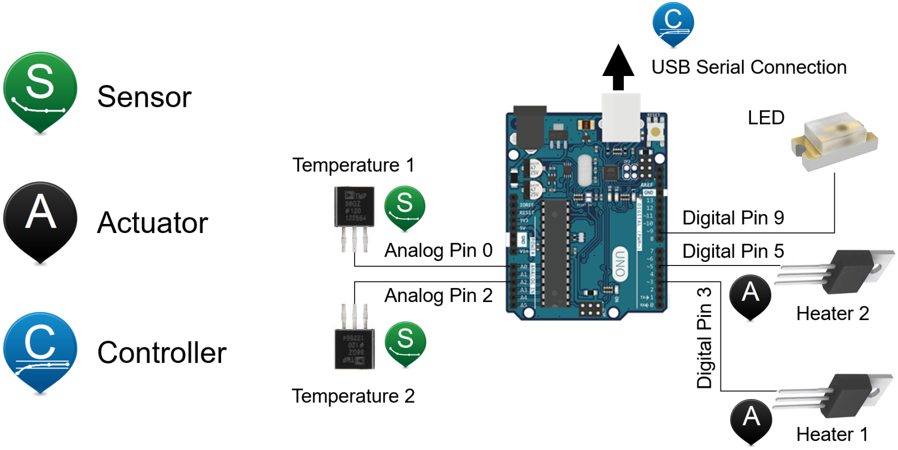
Core Matlab toolboxes

- Control toolbox
- System identification toolbox
- Optimization toolbox
- Statistics and machine learning toolbox

You might need commands from these toolboxes as well

- Signal processing toolbox
- Curve fitting toolbox
- Econometrics toolbox
- Fuzzy logic toolbox

Temperature control laboratory (TCLab)



Link: `https://apmonitor.com/pdc/index.php/Main/ArduinoTemperatureControl`

TCLab model

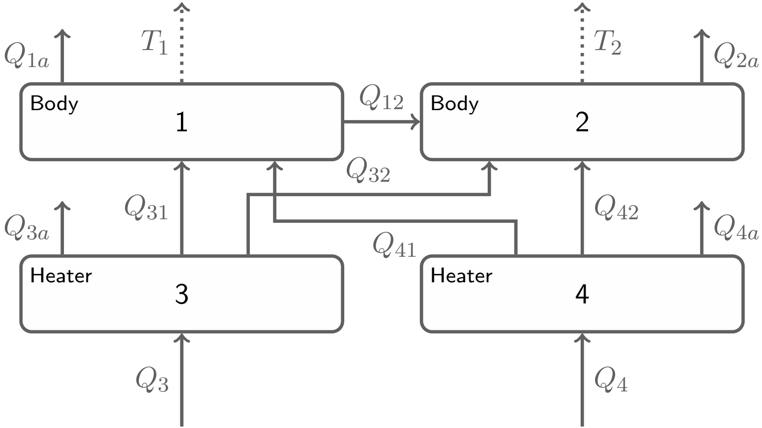


Figure: Four-compartment model of TCLab device.

Demonstration

Today’s Agenda

Probability theory and ARX models

- Probability theory
- Auto-regressive (AR) models
- Auto-regressive models with exogenous inputs (ARX)
- Prediction with ARX models

Probability theory

Stochastic scalar variables

Stochastic variable

$$X \sim \mathcal{F}(\mathbf{p}) \tag{1}$$

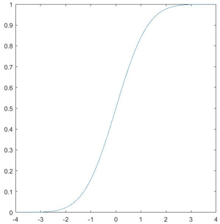
Cumulative distribution function $F_X(y)$ (cdf)

$$F_X(y) = \Pr\{X \leq y\} \in [0, 1], \quad \Pr\{a \leq X \leq b\} = F_X(b) - F_X(a) \tag{2}$$

Probability density function $f_X(z) \geq 0$ (pdf)

$$F_X(y) = \int_{-\infty}^{y} f_X(z)\,\mathrm{d}z, \tag{3}$$

$$F(-\infty) = 0, \quad F(\infty) = 1 \tag{4}$$



Confidence Interval

$1-p$ confidence interval $CI(p)$

$$\Pr\{a \le X \le b\} = 1-p \tag{5}$$

Confidence interval based on inverse cdf

$$\Pr\{X \le a\} = p/2 \quad \text{or} \quad \Pr\{X \le b\} = 1-p/2 \tag{6}$$

$$CI(p) = [F_X^{-1}(p/2), F_X^{-1}(1-p/2)] \tag{7}$$

Use Matlab routines (or look-up tables) to compute $F^{-1}(p/2)$

$$X \in m_X \pm \sigma_X F^{-1}(p/2) \tag{8}$$

Example: Let $X \sim N(10,4)$. Then, a 95% CI is

$$10 - 2\cdot 1.96 \le X \le 10 + 2\cdot 1.96 \text{ or } 6.08 \le X \le 13.92 \tag{9}$$

Moments

For a random variable X

$$\textbf{Nth moment: } \mathbb{E}[X^n] = \int_\Omega x^n f(x)\,\mathrm{d}x \tag{10}$$

Moments represent certain properties of stochastic variables.

$$\textbf{Mean (1st moment): } \mathbb{E}[X] = m_x \tag{11}$$

$$\begin{aligned}\textbf{Variance (2nd central moment): } \mathrm{Var}(X) &= \mathbb{E}[(X - m_x)^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma_x^2 \end{aligned} \tag{12}$$

Moments - Sample Moments

Let $\{x_i\}_{i=1}^{N}$ be samples of X

Estimates of first and second order moments

$$\mathbb{E}[X] \approx \sum_{i=1}^{N} \frac{x_i}{N} \tag{13}$$

$$\mathrm{Var}(X) \approx \sum_{i=1}^{N} \frac{(x_i - \mathbb{E}[X])^2}{N} \tag{14}$$

Unbiased estimate of variance

$$\mathrm{Var}(X) \approx \sum_{i=1}^{N} \frac{(x_i - \mathbb{E}[X])^2}{N-1} \tag{15}$$

$$\tag{16}$$

Probabilities: Joint probability and independence

Marginal probability that a single statement $(X \leq x)$ is true

$$\Pr\{X \leq x\} = F_X(x) \tag{17}$$

Joint probability that two (or more) statements are true

$$\Pr\{X \leq x, Y \leq y\} = F_{X,Y}(x, y) \tag{18}$$

Compute the **marginal distribution** from the joint distribution

$$f_X(x) = \int_{\Omega_y} f_{X,Y}(x, y)\, \mathrm{d}y \tag{19}$$

Joint distributions for **independent** variables

$$F_{X,Y}(x, y) = F_X(x) F_Y(y), \quad f_{X,Y}(x, y) = f_X(x) f_Y(y) \tag{20}$$

Covariance

Covariance is a measure of how two stochastic variables varies relatively to each other

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - m_x)(Y - m_y)] \tag{21}$$

Variance is covariance between the same variable

$$\mathrm{Var}(X) = \mathrm{Cov}(X, X) \tag{22}$$

Correlation coefficient

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}, \quad -1 \leq \rho \leq 1 \tag{23}$$

Covariance of independent variables

$$\mathrm{Cov}(X, Y) = \rho = 0 \tag{24}$$

Note: The reverse if not true

Conditional distribution – Bayes' theorem

Conditional probability and Bayes' theorem

$$\Pr(A|B)\Pr(B) = \Pr(A,B) = \Pr(B|A)\Pr(A), \tag{25}$$

$$\Pr\{X \le x|Y \le y\}\Pr\{Y \le y\} = \Pr\{X \le x, Y \le y\} \tag{26}$$

$$= \Pr\{Y \le y|X \le x\}\Pr\{X \le x\} \tag{27}$$

Conditional probability density function

$$f_{X|Y}(x|y)f_Y(y) = f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) \tag{28}$$

The same can be done for the moments if $\mathrm{Var}(X|Y) < \infty$ exists

$$\mathbb{E}[X|Y] = m_{x|y} = \int_{\Omega_x} x f_{X|Y}(x|y)\,\mathrm{d}x \tag{29}$$

$$\mathrm{Var}(X|Y) = \mathbb{E}[(X - m_{x|y})^2|Y] \tag{30}$$

Different Distributions - Gaussian and χ^2

Gaussian/normal distribution

$$X \sim N(m_x, \sigma_x^2) \tag{31}$$

$$Y = \frac{X - m_x}{\sigma_x} \sim N(0,1) \quad \text{standard Gaussian} \tag{32}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x-m_x)^2}{2\sigma_x^2}\right) \tag{33}$$

$$F_X(x) = F_Y\left(\frac{x - m_x}{\sigma_x}\right) \tag{34}$$

χ^2-distribution

$$X = \sum_{i=1}^{n} \psi_i^2 \sim \chi^2(n), \quad \psi_i \sim N(0,1), \quad \psi_i \perp \psi_j \tag{35}$$

$$f(x) = \frac{1}{\Gamma(n/2)} x^{n/2-1} \exp\left(-\frac{x}{2}\right) \tag{36}$$

$$\mathbb{E}[X] = n \quad \mathrm{Var}(X) = 2n \tag{37}$$

Different Distributions - Gamma

Gamma distribution ($\chi^2(n) = \Gamma(n/2, 2)$)

$$X \sim \Gamma(k, \theta), \quad 0 < X < \infty \tag{38}$$

$$f_X(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right) \tag{39}$$

$$\mathbb{E}[X] = k\theta, \quad \mathrm{Var}(X) = k\theta^2 \tag{40}$$

Gamma function

$$\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt \tag{41}$$

$$\Gamma(k+1) = k\Gamma(k) \tag{42}$$

$$\Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{43}$$

Erlang distribution (Gamma distribution for integer values of k)

$$\Gamma(k) = (k-1)!, \quad \Gamma\left(k + \frac{1}{2}\right) = \frac{(2k-1)!}{2^k}\sqrt{\pi} \tag{44}$$

Other Related Distributions

The F-distribution

$$X = \frac{Zm}{Yn} \sim F(n, m) \tag{45}$$

$$Z \sim \chi^2(n), \quad Y \sim \chi^2(m), \quad Z \perp Y \tag{46}$$

Student's t-distribution

$$X = \frac{Z}{\sqrt{Y}}\sqrt{n} \sim t(n) \tag{47}$$

$$Z \in N(0, 1), \quad Y \sim \chi^2(n), \quad Z \perp Y \tag{48}$$

The Rayleigh distribution:

$$X = \sqrt{Y_1^2 + Y_2^2} \sim Ray(\sigma_y^2), \quad Y_i \sim N_{iid}(0, \sigma_y^2) \tag{49}$$

Stochastics in Matlab

Sampling from distributions in Matlab

Stochastic Processes

A stochastic process can be described using a marginal cdf or pdf

$$F_{X_t}(x_t, t) = \Pr\{X_t \leq x_t\} \tag{50}$$

$$f_{X_t}(x_t, t) = \nabla_{x_t} F_{X_t}(x_t, t) \tag{51}$$

or if the different times are related, using joint probabilities

$$F_{X_t, X_s}(x_t, x_s, t, s) = \Pr\{X_t \leq x_t, X_s \leq x_s\} \tag{52}$$

Properties of Stochastic processes

Mean

$$m_x(t) = \mathbb{E}[x(t)] = \int_{-\infty}^{\infty} z f_{x(t)}(z)\,\mathrm{d}z, \tag{53}$$

Variance

$$P_x(t) = \mathrm{Var}(x(t)) = \mathbb{E}\Big[\big(x(t) - \mathbb{E}[x(t)]\big)\big(x(t) - \mathbb{E}[x(t)]\big)^T\Big], \tag{54}$$

Auto-covariance

$$r_x(t_1, t_2) = \mathrm{Cov}(x(t_1), x(t_2)) = \mathbb{E}\Big[\big(x(t_1) - \mathbb{E}[x(t_1)]\big)\big(x(t_2) - \mathbb{E}[x(t_2)]\big)^T\Big], \tag{55}$$

Note that $r_x(t,t) = P_x(t)$

Auto-covariance and auto-correlation

Auto-correlation function

$$\rho_x(t_1, t_2) = \frac{r_x(t_1, t_2)}{\sqrt{P_x(t_1)P_x(t_2)}} \tag{56}$$

For stationary processes, only the time difference $\tau = t_1 - t_2$ is relevant

$$r_x(\tau) = \text{Cov}\left[x(t), x(t+\tau)\right] \tag{57a}$$

$$\rho_x(\tau) = \frac{r_x(\tau)}{P_x(\tau)} \tag{57b}$$

Stationarity

Steady state of deterministic systems: The state does not change in time

Stationary distributions of stochastic systems: The state distribution does not change in time

Strong stationarity

$$f_{x(t_1),\ldots,x(t_n)}(z_1,\ldots,z_n) = f_{x(t_1+\Delta t),\ldots,x(t_n+\Delta t)}(z_1,\ldots,z_n), \tag{58}$$

for any $n \in \mathbb{N}$ and $\Delta t \in \mathbb{R}$.

Weak stationarity: The first two moments (mean and covariance) do not change in time

Normal process and Markov process

Normal process: Any probability density function $f_{x(t_1),\ldots,x(t_n)}(z_1,\ldots,z_n)$ is a multivariate normal distribution for any $n \in \mathbb{N}$

Probability density function with mean μ and covariance Σ

$$f_Y(y) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right) \tag{59}$$

Markov process: For any $t_1 < t_2 < \cdots < t_n$, the distribution of $x(t_n)$ given $(x(t_1),\ldots,x(t_{n-1}))$ is the same as the distribution of $x(t_n)$ given $x(t_{n-1})$

$$\Pr\Big(x(t_n) \leq x \mid x(t_{n-1}),\ldots,x(t_1)\Big) = \Pr\Big(x(t_n) \leq x \mid x(t_{n-1})\Big) \tag{60}$$

Auto-regressive models

The Auto-Regressive (AR) Process

AR(m) process

$$y_t + \sum_{k=1}^{m} a_k y_{t-k} = e_t, \qquad a_0 = 1 \tag{61}$$

$\{e_t\}$ is a white-noise process

The Auto-Regressive (AR) Process

AR(m) process (compact notation)

$$A(q^{-1})y_t = e_t \tag{62}$$

Polynomial

$$A(q^{-1}) = 1 + \sum_{k=1}^{m} a_k q^{-k} \tag{63}$$

It is called *auto-regressive* because y_t can be viewed as a regression on past values

$$y_t = e_t - \sum_{k=1}^{m} a_k y_{t-k} \tag{64}$$

The Auto-Regressive (AR) Process - stability

An AR(m) process is stable if the roots, λ, of the characteristic equation

$$a_0 + a_1\lambda^{-1} + \cdots + a_m\lambda^{-m} = 0 \tag{65}$$

lie within the unit circle

Multiply both sides by λ^m (assuming λ is nonzero)

$$a_0\lambda^m + a_1\lambda^{m-1} + \cdots + a_m = 0 \tag{66}$$

Hint: If the Matlab vector `A` contains the elements a_0, a_1, ..., a_m (in that order), you can use the command `roots(A)` to find the poles (double-check with the documentation)

The Auto-Regressive (AR) Process – Mean

Mean

$$A(q^{-1})\mathbb{E}[y_t] = \mathbb{E}[e_t] = 0 \tag{67}$$

Stationary mean, $\mathbb{E}[y_t] = \bar{y}$ for all t

$$A(q^{-1})\mathbb{E}[y_t] = A(q^{-1})\bar{y} = A(0)\bar{y} = 0 \tag{68}$$

$A(q^{-1})$ applied to a constant (e.g., $\bar{y}$) is just the sum of the coefficients

$$A(0) = \sum_{k=0}^{m} a_k \tag{69}$$

If $A(0) \neq 0$

$$\bar{y} = 0 \tag{70}$$

The Auto-Regressive (AR) Process – Auto-covariance

Auto-covariance function of an AR(m) process

$$\gamma(k) = \mathrm{Cov}(y_t, y_{t-k}) \tag{71}$$

Auto-covariance

$$\gamma(k) + \sum_{j=1}^{m} a_j \gamma(k-j) = 0, \qquad k > 0 \tag{72}$$

Initial condition

$$\gamma(0) + \sum_{j=1}^{m} a_j \gamma(j) = \sigma_e^2 \tag{73}$$

Symmetry of auto-covariance functions: $\gamma(k) = \gamma(-k)$

The Auto-Regressive Process with Exogenous inputs (ARX)

ARX process

$$y_t + \sum_{k=1}^{m} a_k y_{t-k} = \sum_{k=1}^{n} b_k u_{t-k} + e_t, \qquad a_0 = 1 \qquad (74)$$

$\{e_t\}$ is a white-noise process

ARX process (using polynomials)

$$A(q^{-1}) y_t = B(q^{-1}) u_t + e_t \qquad (75)$$

The Auto-Regressive Process with Exogenous inputs (ARX)

Mean

$$A(q^{-1})\mathbb{E}[y_t] = B(q^{-1})\mathbb{E}[u_t] + \mathbb{E}[e_t] = 0 \tag{76}$$

Stationary mean, $\mathbb{E}[y_t] = \bar{y}$ and $\mathbb{E}[u_t] = u_t = \bar{u}$ for all t

$$A(q^{-1})\mathbb{E}[y_t] = A(1)\bar{y} = B(q^{-1})\mathbb{E}[u_t] = B(1)\bar{u} \tag{77}$$

If $A(1) \neq 0$

$$\bar{y} = \frac{B(1)}{A(1)}\bar{u} \tag{78}$$

Diophantine equations

Polynomials and Transfer functions

Polynomials

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_n q^{-n} \tag{79}$$

The polynomial is order n if $b_n \neq 0$ and $b_i = 0$ for $i > n$

If $b_0 = 1$, the polynomial is *monic*

A transfer function $H(q)$ can be written in infinitely many ways

$$H(q) = \frac{B(q^{-1})}{A(q^{-1})} = \frac{C(q^{-1})B(q^{-1})}{C(q^{-1})A(q^{-1})} \tag{80}$$

Polynomials and Transfer functions

$$\frac{B(q^{-1})}{A(q^{-1})} = \frac{b_0 + b_1 q^{-1} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} \tag{81}$$

$$= b_0 - b_0 + \frac{b_0 + b_1 q^{-1} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} \tag{82}$$

$$= b_0 - \frac{b_0(1 + a_1 q^{-1} + \cdots + a_n q^{-n})}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} + \frac{b_0 + b_1 q^{-1} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} \tag{83}$$

$$= b_0 + q^{-1} \frac{(b_1 - b_0 a_1) + (b_2 - b_0 a_2) q^{-1} + \cdots + (b_n - b_0 a_n) q^{-(n-1)}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} \tag{84}$$

Define the transfer function

$$H(q) = \frac{B(q^{-1})}{A(q^{-1})} = g_0 + q^{-1} \frac{S_1(q^{-1})}{A(q^{-1})}, \tag{85}$$

$$S_1(q^{-1}) = s_0 + s_1 q^{-1} + \ldots + s_{n_1} q^{-n_1}, \tag{86}$$

$$g_0 = b_0, \quad s_i = b_{i+1} - b_0 a_{i+1} \tag{87}$$

$n_1 = n - 1$ is the order of S_1

Polynomials and Transfer functions

Repeat the rewriting for $\frac{S_1}{A}$, $\frac{S_2}{A}$, etc.

$$H(q) = \frac{B(q)}{A(q)} = g_0 + g_1 q^{-1} + \cdots + g_{m-1} q^{-(m-1)} + q^{-m} \frac{S_m(q^{-1})}{A(q^{-1})} \quad (88)$$

$$= G_m(q^{-1}) + q^{-m} \frac{S_m(q^{-1})}{A(q^{-1})} \quad (89)$$

Diophantine equation

$$B(q^{-1}) = A(q^{-1}) G_m(q^{-1}) + q^{-m} S_m(q^{-1}) \quad (90)$$

The order of S_m is $\max(n_a - 1, n_b - m)$ and the order of G_m is $m - 1$

Solving simple Diophantine equations

This (simple) Diophantine equation can be solved iteratively

```
% Initialize
G = [];
S = [B, 0]; % Pad B with zeros to make S as long as A

for i = 1:m
    % Augment with first element of S
    G = [G, S(1)];

    % Update S
    S = [S(2:end) - S(1)*A(2:end), 0];
end

% Remove last element
S = S(1:end-1);
```

Prediction

Prediction - AR process

Weakly stationary process

$$A(q^{-1})y_t = e_t \tag{91}$$

e_t is a white noise signal $\mathcal{F}(0, \sigma^2)$ and A is monic

m-step prediction based on solution to the Diophantine equation

$$y_{t+m} = \frac{1}{A(q^{-1})}e_{t+m} = G_m(q^{-1})e_{t+m} + \frac{S_m(q^{-1})}{A(q^{-1})}e_t \tag{92}$$

Prediction and error

$$\hat{y}_{t+m|t} = \frac{S_m(q^{-1})}{A(q^{-1})}e_t = \frac{S_m(q^{-1})}{A(q^{-1})}A(q^{-1})y_t = S_m(q^{-1})y_t, \tag{93}$$

$$\tilde{y}_{t+m|t} = G_m(q^{-1})e_{t+m} \tag{94}$$

$\hat{y}_t$ and $\tilde{y}_t$ are independent

Prediction - ARX process

System

$$A(q^{-1})y_t = q^{-k}B(q^{-1})u_t + e_t \tag{95}$$

k is the control delay

m-step prediction

$$\hat{y}_{t+m|t} = B(q^{-1})G_m(q^{-1})u_{t+m-k} + S_m(q^{-1})y_t, \tag{96}$$

$$\tilde{y}_{t+m|t} = G_m(q^{-1})e_{t+m} \tag{97}$$

Diophantine equation

$$1 = A(q^{-1})G_m(q^{-1}) + q^{-m}S_m(q^{-1}) \tag{98}$$

The order of G and S are $m-1$ and $\max(n_a - 1, 1 - m)$ and $G(0) = 1$

Proof of ARX prediction

Rewrite future output using the Diophantine equation

$$y_{t+m} = \left(A(q^{-1})G_m(q^{-1}) + q^{-m}S_m(q^{-1})\right) y_{t+m} \quad (99)$$

$$= G_m(q^{-1})A(q^{-1})y_{t+m} + S_m(q^{-1})y_t \quad (100)$$

Substitute system description

$$y_{t+m} = G_m(q^{-1})(B(q^{-1})u_{t+m-k} + e_{t+m}) + S_m(q^{-1})y_t \quad (101)$$

$$= G_m(q^{-1})B(q^{-1})u_{t+m-k} + S_m(q^{-1})y_t + G_m(q^{-1})e_{t+m} \quad (102)$$

$$= \hat{y}_{t+m|t} + \tilde{y}_{t+m|t} \quad (103)$$

The prediction $\hat{y}_{t+m|t}$ is uncorrelated with the prediction error $\tilde{y}_{t+m|t}$ because y_t is independent of the future noises, e_t (recall that G_m is of order $m-1$ so the error term involves $e_{t+1}, \ldots, e_{t+m}$)

Questions?