Filtering and Identification

Day 3 - Lecture 2:
The optimal predictor for input-output (i/o) models and Prediction Error Optimization

Michel Verhaegen

Division of Automatic Control - EE

Overview

- **Recap “goals” of the course**
- The Output-error Identification “recipe”
- The prediction-error Identification “recipe”
- Case Study: the acoustical duct

Recap “goal” of the course

- Estimate linear static models from data:

$$\min_x \|y - Fx\|_W^2$$

  Relevant and to introduce the numerical fundaments of this course!

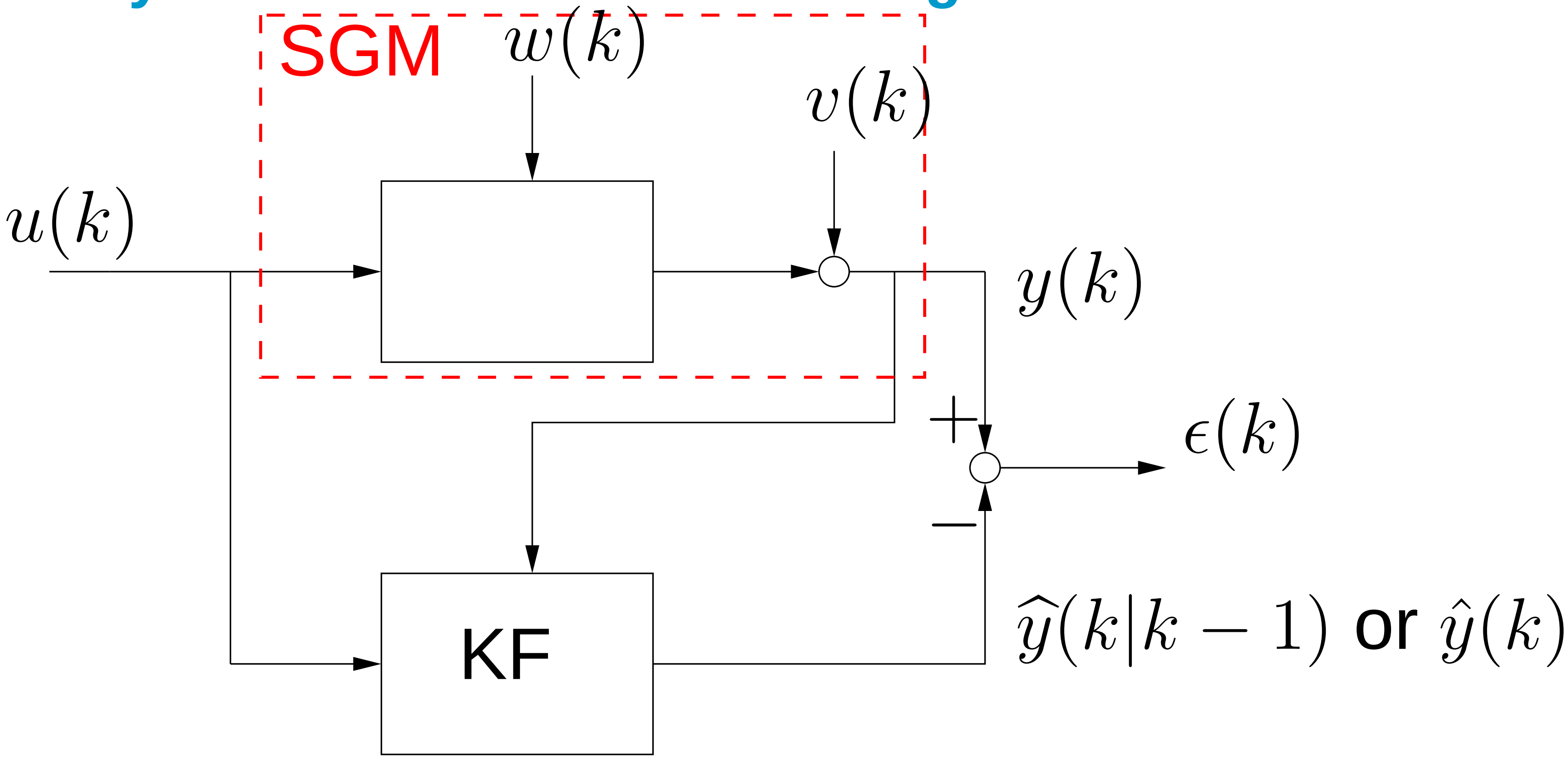
- Estimate the state of an LTI state space model:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ke(k) \quad e(k) \sim (0, R_e) \\ y(k) &= Cx(k) + Du(k) + e(k) \end{aligned}$$

  given the model (A, B, C, D) and K and input-output data.

- “Identify” state space model matrices (A, B, C, D, K) from input-output data $\{u(k), y(k)\}_{k=1}^N$.

The system identification challenge



1. Generally SGM is not known!
2. Then we have an “identification problem”
3. First we study the challenges with “classical” PEM

The “classical” identification “recipe”

1. Assume a SGM — say $\widehat{\mathrm{SGM}}(\theta)$
2. Derive from $\widehat{\mathrm{SGM}}(\theta)$ the “optimal” prediction of the output — denoted as $\hat{y}(k|k-1;\theta)$ (in short $\hat{y}(k,\theta)$)
3. Find the “best” estimate $\hat{\theta}$ by minimizing the cost function:

$$\frac{1}{N}\sum_{k=1}^{N}(y(k)-\hat{y}(k,\theta))^2$$

w.r.t. θ.

Overview

- Recap “goals” of the course
- **The Output-error Identification “recipe”**
- The prediction-error Identification “recipe”
- Case Study: the acoustical duct

The OE Identification recipe

1. Given the "true" SGM:

   $$y(k) = G(q)u(k) + v(k) \in \mathbb{R} \quad E[u(k)v(\ell)^T] = 0$$

2. Model $\widehat{\text{SGM}}(\theta)$: $y(k,\theta) = \underbrace{\frac{B(q,b_i)}{A(q,a_i)}u(k)}_{z(k)} + e(k) \quad e(k) \sim (0, R)$

   with $B(q,b_i) = b_1 q^{\boxed{-n_k}} + \cdots + b_{n_b} q^{\boxed{-n_b - n_k + 1}}$ and $A(q,a_i) = \boxed{1} + a_1 q^{-1} + \cdots + a_{n_a} q^{\boxed{-n_a}}$ ($\boxed{\text{monic}}$) and q is the time shift operator.

   Generally $n_k = 1$ and $n_a = n_b$.

with $\theta = \Big[a_i, b_i, \cdots\Big]$.

Example OE model 3rd-order:

Consider,

$$y(k) = \underbrace{\frac{b_1 q^{-1} + \cdots + b_3 q^{-3}}{1 + a_1 q^{-1} + \cdots + a_3 q^{-3}} u(k)}_{z(k)} + e(k)$$

Then this is explicitly denoted as,

$$\begin{cases} z(k) + a_1 z(k-1) + a_2 z(k-2) + a_3 z(k-3) & = & b_1 u(k-1) + \\ & & + b_2 u(k-2) + b_3 u(k-3) \\ y(k) & = & z(k) + e(k) \end{cases}$$

Finding "optimal" predictor using SS innovation form

Example OE model 3rd-order (c'td):

$$z(k) = \underbrace{-a_1 z(k-1) + b_1 u(k-1) \underbrace{-a_2 z(k-2) + b_2 u(k-2) \underbrace{-a_3 z(k-3) + b_3 u(k-3)}_{x_3(k-2)}}_{x_2(k-1)}}_{x_1(k)}$$

The definition of the states yields the following SS model of the OE $\widehat{\mathrm{SGM}}(\theta)$

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \underbrace{\begin{bmatrix} -a_1 & 1 & 0 \\ -a_2 & 0 & 1 \\ -a_3 & 0 & 0 \end{bmatrix}}_{A_3(a)} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{B_3(b)} u(k)$$

$$y(k) = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{C_3} x(k) + e(k)$$

What is the optimal predictor of the output?

The “optimal” predictor of an output-error model

The innovation form of the OE-SGM model reads for $\theta^T = \begin{bmatrix} a^T & b^T & x(0)^T & vec(R)^T \end{bmatrix}$:

$$\begin{aligned} x(k+1,\theta) &= A(a)x(k,\theta) + B(b)u(k) \\ y(k,\theta) &= Cx(k,\theta) + e(k) \quad e(k) \sim (0, R) \end{aligned}$$

The Kalman filter analysis directly shows that the stationary optimal predictor equals:

$$\begin{aligned} \hat{x}(k+1,\theta) &= A(a)\hat{x}(k,\theta) + B(b)u(k) \\ \hat{y}(k,\theta) &= C\hat{x}(k,\theta) \end{aligned}$$

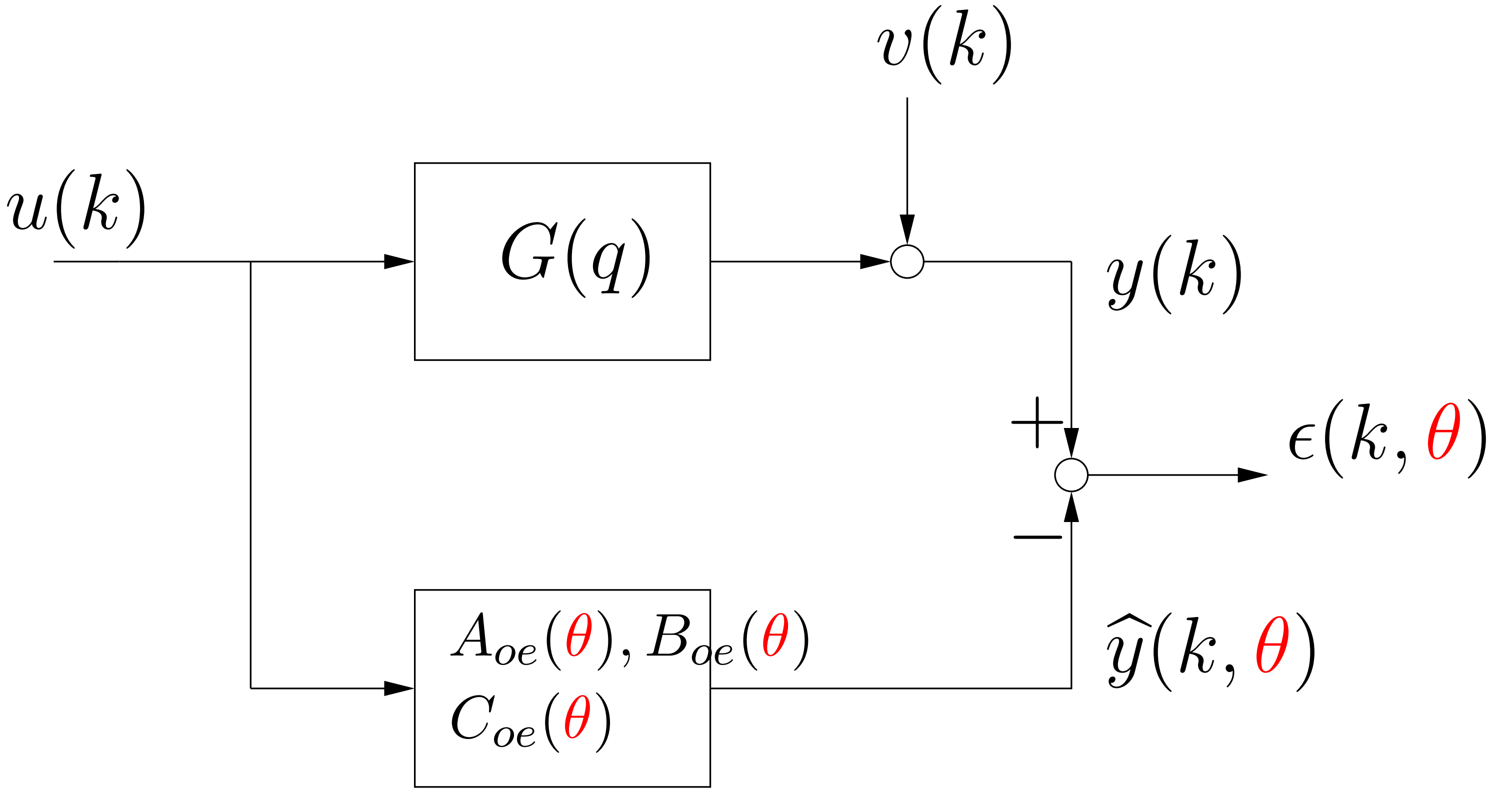
The stationary Kalman filter

The stationary solution for an LTI system is characterized by the DARE, which for the given OE model reads:

$$P = APA^T - APC^T(CPC^T + R)^{-1}CPA^T$$

The solution that satisfies this equation is $P \equiv 0$ and hence the KF gain K is also zero!

Output-error problem



$$\hat{\theta}_N = \text{argmin} J_N(\theta) \quad \text{with} \quad J_N(\theta) = \frac{1}{N}\sum_{k=1}^{N} \|y(k) - \hat{y}(k,\theta)\|_2^2$$

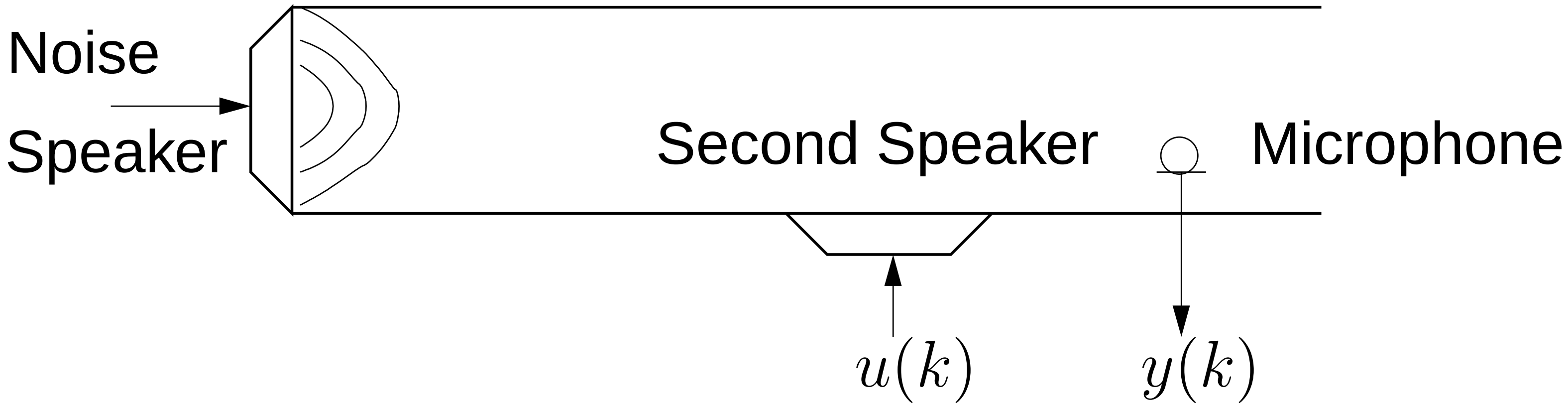
Challenges for the classical OE Identification approach

1. You need to fix the order of the SS model a priori!
2. The parametrization for MIMO systems is a difficult?
3. The optimization problem is non-convex, except for the FIR model.

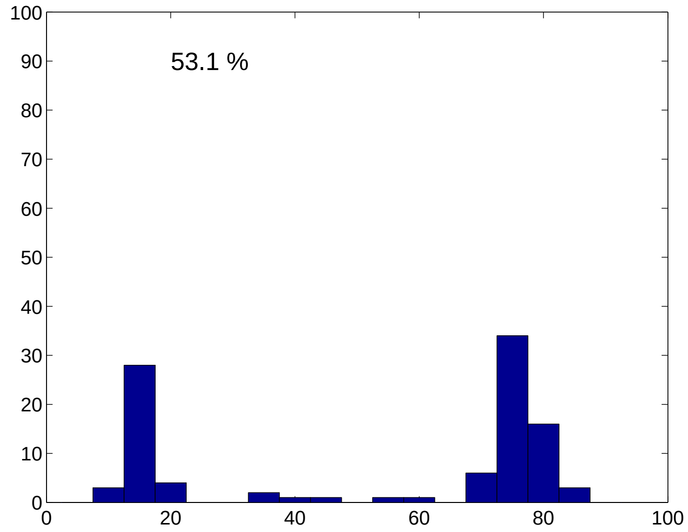
linCost.m, nlCost.m

Consequence of non-convexity cost function

Example: OE model (n=6) for Acoustical duct



Results of a Monte Carlo simulation of 100 runs



Model error in percentage

100% $=$ perfect fit

Overview

- Recap “goals” of the course
- The Output-error Identification “recipe”
- **The prediction-error Identification “recipe”**
- Case Study: the acoustical duct

The PE Identification recipe

1. Given the "true" SGM:

   $y(k) = G(q)u(k) + H(q)e(k) \in \mathbb{R} \quad E[u(k)e(\ell)^T] = 0$ and $e(k) \sim (0, R)$

2. Model $\widehat{\mathrm{SGM}}(\theta)$: many possibilities!

   The AR(MA)X model:

$$y(k, \theta) = \frac{b_1 q^{-1} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} u(k) + \frac{1 + \left(c_1 q^{-1} + \cdots + c_n q^{-n}\right)}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} e(k)$$

with $\theta = \left[a_i, b_i, c_i, \cdots\right]$.

Finding “optimal” predictor using SS innovation form

The AR(MA)X model:

$$y(k,\theta) = \frac{b_1 q^{-1} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} u(k) + \frac{1 + \left(c_1 q^{-1} + \cdots + c_n q^{-n}\right)}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}} e(k)$$

with $\theta = \left[a_i, b_i, c_i, \cdots\right]$. This model parametrizes the state space innovation model as (Lemma 8.3):

$$x(k+1,\theta) = \underbrace{\begin{bmatrix} -a_1 & 1 & & 0 \\ & & \ddots & \\ -a_{n-1} & & & 1 \\ -a_n & & \cdots & 0 \end{bmatrix}}_{A_n(a)} x(k,\theta) + \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_{n-1} \\ a_n \end{bmatrix}}_{B_n(b)} u(k) + \underbrace{\begin{bmatrix} (c_1) - a_1 \\ \vdots \\ \\ (c_n) - a_n \end{bmatrix}}_{K_n(k)} e(k)$$

$$y(k,\theta) = \underbrace{\begin{bmatrix} 1 & \cdots & 0 \end{bmatrix}}_{C_n} x(k,\theta) + e(k)$$

What is the optimal predictor of the output?

The "optimal" predictor of an AR(MA)X model

The "optimal" predictor derived from the innovation form equals:

$$\hat{x}(k+1,\theta) = \begin{bmatrix} -c_1 & 1 & & 0 \\ & & \ddots & \\ -c_{n-1} & & & 1 \\ -c_n & & \cdots & 0 \end{bmatrix} \hat{x}(k,\theta) + \begin{bmatrix} b_1 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} u(k) + \begin{bmatrix} k_1 \\ \vdots \\ k_{n-1} \\ k_n \end{bmatrix} y(k)$$

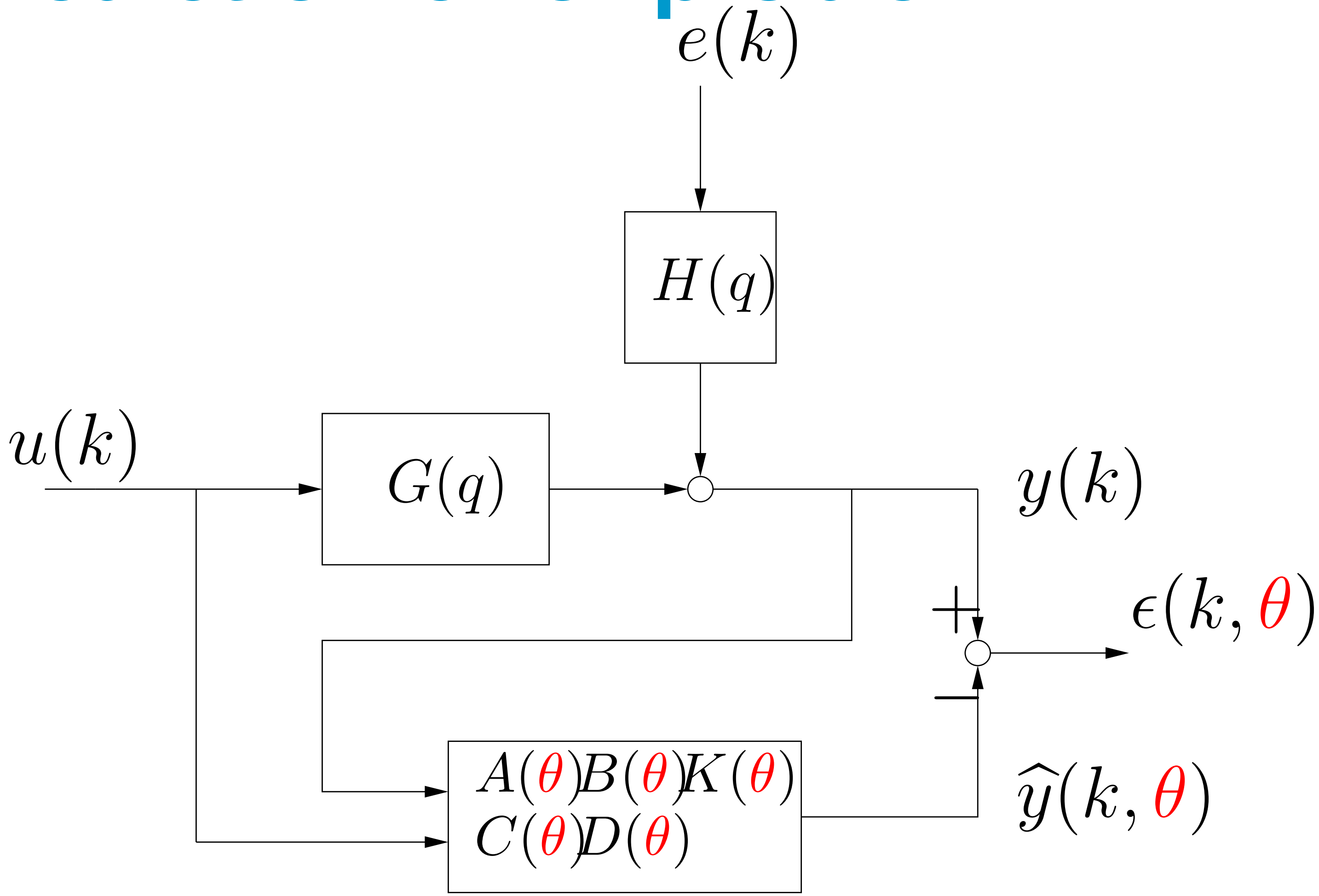
$$\hat{y}(k,\theta) = \begin{bmatrix} 1 & \cdots & 0 \end{bmatrix} \hat{x}(k,\theta)$$

with $k_1 = (c_i) - a_i$. This is given in transfer function form as:

$$\boxed{\hat{y}(k,\theta) = \frac{B(q,\theta)}{\left(C(q,\theta)\right)} u(k) + \frac{\left(C(q,\theta)\right) - A(q,\theta)}{\left(C(q,\theta)\right)} y(k)}$$

The ARX predictor is linear in the parameters a_i, b_i

Prediction-error problem

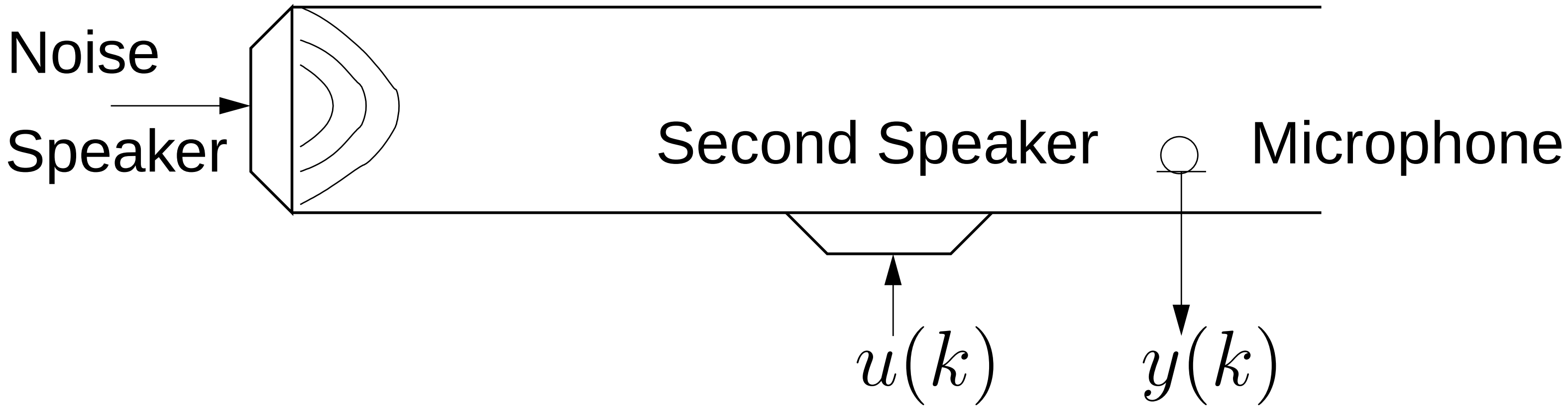


$$\hat{\theta}_N = \text{argmin} J_N(\theta) \quad \text{with} \quad J_N(\theta) = \frac{1}{N}\sum_{k=1}^{N} \|y(k) - \hat{y}(k,\theta)\|_2^2$$

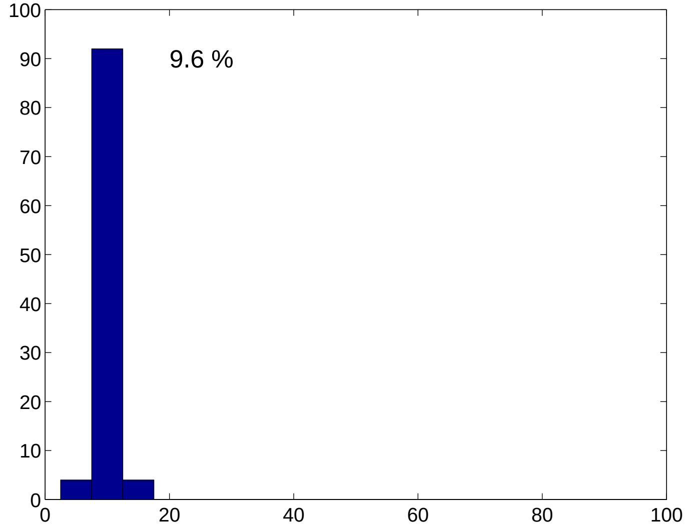
Challenges for the classical PEM Identification approach

1. You need to fix the order of the SS model a priori!
2. The parametrization for MIMO systems is a difficult? and many possibilities to model the stochastic part!
3. The optimization problem is non-convex, except for the ARX model.

Example: ARX model (n=6) for Acoustical duct



Results of a Monte Carlo simulation of 100 runs



Model error in percentage
100% $=$ perfect fit

Preparation for this afternoon

Preparation:

1. Study Chapter 5 (5.5.2 - 5.8) for the KF problem.
2. Study Chapter 8 (8.3) for the predictor parametrization.

Download Homework 3