System Identification
Lecture 5

PE Method - Consistency Analysis

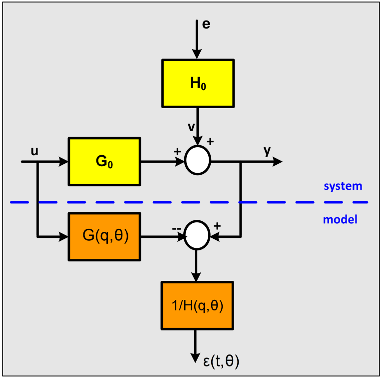
Paul Van den Hof

Control Systems Group
Department of Electrical Engineering
Eindhoven University of Technology

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Contents

The motivating question:
Do the choices that we have made so far (predictor,model set,quadratic identification criterion) lead to attractive identification results?



Prediction error:

$$\varepsilon(t,\theta) = H(q,\theta)^{-1}[y(t) - G(q,\theta)u(t)]$$

Model set:

$$\mathcal{M} = \{(G(q,\theta), H(q,\theta)), \theta \in \Theta \subset \mathbb{R}^d\}$$

Identification criterion:

$$\hat{\theta}_N = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t,\theta)$$

What is an attractive identification result?

Minimum requirement:
If the data has been generated by a system (G_0, H_0) then we should be able to estimate this correctly (consistently) on the basis of input/output data, provided that we choose a large enough model set.

In this lecture we are going to verify under which conditions this property holds.

Two-steps analysis:

- Consequence of approximating

$$\bar{\mathbb{E}}\varepsilon^2(t,\theta) \quad \text{by} \quad \underbrace{\frac{1}{N}\sum_{t=0}^{N-1}\varepsilon^2(t,\theta)}_{V_N(\theta,Z^N)}$$

  (convergence)
- Consistency: will the estimated models be correct (w.p.1), if we choose an appropriate model set?

Contents

- Asymptotic convergence
- Additional conditions / notions:
    - Data generating system
    - System in the model set
    - Persistently exciting input signals
- Consistency
- Validation for consistency

Assumptions for analysis:

- Data has been generated by a linear dynamical system $\mathcal{S}$, possibly operating in a (stabilized) feedback situation, with bounded excitation signals, and v having bounded moments of order 4;
- $\mathcal{M}$ is uniformly stable ($\rightarrow$ the predictor filters and their derivatives w.r.t. θ are uniformly stable).

Convergence result

Under the above assumptions, $\hat{\theta}_N = \arg\min_\theta V_N(\theta, Z^N)$, for $N \rightarrow \infty$ converges with probability 1 to the minimizing argument θ^* of

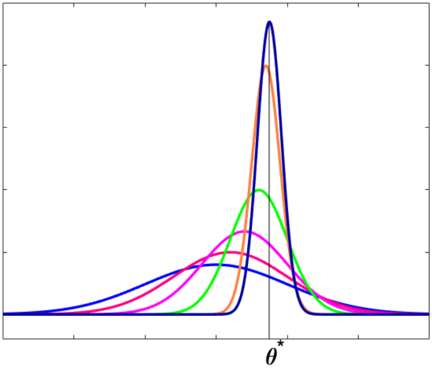
$$\bar{V}(\theta) := \bar{\mathbb{E}}\varepsilon^2(t,\theta)$$

Differently stated:

$$\hat{\theta}_N \to \theta^* := \arg\min_{\theta} \bar{\mathbb{E}}\varepsilon^2(t,\theta) \quad \text{w.p. } 1 \text{ for } N \to \infty.$$

The estimate converges to the best possible predictor (in the sense of $\bar{\mathbb{E}}\varepsilon^2$) within the model set $\mathcal{M}$, and becomes independent of the particular noise sequence (variance goes to 0)

Probability density function of $\hat{\theta}_N$ for increasing values of N:



For $N \to \infty$, the pdf of $\hat{\theta}_N$ converges to a Dirac impulse at θ^*.

Parameter estimate asymptotically (in N) converges to the limit point θ^* (with probability one).

Next question:

Under which conditions does θ^* represent the correct model?

Example 4.8.3 (Calculate θ^* in a particular situation)

ARMAX-system:

$$y(t) = \frac{b_0 q^{-1}}{1 + a_0 q^{-1}} u(t) + \frac{1 + c_0 q^{-1}}{1 + a_0 q^{-1}} e(t)$$

$$y(t) + a_0 y(t-1) = b_0 u(t-1) + e(t) + c_0 e(t-1)$$

u, e white, unit variance, and uncorrelated with eachother.

ARX-model:

$$y(t) = \frac{b q^{-1}}{1 + a q^{-1}} u(t) + \frac{1}{1 + a q^{-1}} \varepsilon(t, \theta)$$

$$\varepsilon(t,\theta) = y(t) + a y(t-1) - b u(t-1) \qquad \theta = [a \quad b]^T$$

Note: $\mathcal{S} \notin \mathcal{M}$, $G_0 \in \mathcal{G}$.

$$\varepsilon(t,\theta) = y(t) + ay(t-1) - bu(t-1)$$

$$\bar{V}(\theta) = (1+a^2)R_y(0) + 2aR_y(1) + b^2 - 2bR_{yu}(1) - 2abR_{yu}(0)$$

For finding the values of the correlations: take system equation

$$y(t) + a_0 y(t-1) = b_0 u(t-1) + e(t) + c_0 e(t-1)$$

and postmultiply with respectively $u(t)$, $u(t-1)$, $e(t)$ and $y(t-1)$ and taking $\bar{\mathbb{E}}$ to obtain:

$$R_{yu}(0) + a_0 R_{yu}(-1) = b_0 R_u(1)$$

$$R_{yu}(1) + a_0 R_{yu}(0) = b_0 R_u(0)$$

$$R_{ye}(0) + a_0 R_{ye}(-1) = R_e(0) + c_0 R_e(1)$$

$$R_y(1) + a_0 R_y(0) = b_0 R_{yu}(0) + c_0 R_{ye}(0).$$

Since $R_{yu}(-1) = R_u(1) = 0$ it follows that $R_{yu}(0) = 0$.
Then $R_{yu}(1) = b_0$.
With $R_{ye}(-1) = R_e(1) = 0$ it follows that $R_{ye}(0) = 1$
Then $R_y(1) + a_0 R_y(0) = c_0$.

Leading to:

$$\begin{aligned}\bar{V}(\theta) &= \bar{\mathbb{E}}[y(t) + ay(t-1) - bu(t-1)]^2 \\ &= [1 + a^2 - 2aa_0]r_0 + 2ac_0 + b^2 - 2bb_0\end{aligned}$$

with $r_0 = R_y(0)$.

Since $\bar{V}(\theta)$ is quadratic in a and b, minimization of $\bar{V}(\theta)$ is obtained by:

$$\left.\frac{\partial \bar{V}(\theta)}{\partial a}\right|_{a=a^*} = 0 \quad \Leftrightarrow \quad 2(a^* - a_0)r_0 + 2c_0 = 0$$

$$\left.\frac{\partial \bar{V}(\theta)}{\partial b}\right|_{b=b^*} = 0 \quad \Leftrightarrow \quad 2(b^* - b_0) = 0$$

leading to $b^* = b_0$ en $a^* = a_0 - c_0/r_0$.

Note:

- The parameter a is estimated “incorrectly”, but such that the prediction error is minimal in a quadratic sense.
- The correct value is obtained when $c_0 = 0$, i.e. when the system fits into an ARX structure.

Persistence of excitation

Definition - persistently exciting input

A quasi-stationary signal u is persistently exciting of order n if the spectral density $\Phi_u(\omega)$ is unequal to 0 in n points in the interval $(-\pi, \pi]$.

Example

The signal

$$u(t) = sin(\omega_0 t)$$

is persistently exciting of order 2.
(Φ_u has a contribution in $\pm\omega_0$)
2 degrees of freedom (amplitude and phase).

Equivalent formulation

Proposition

A quasi-stationary signal u is persistently exciting of order n if and only if the (Toeplitz) matrix

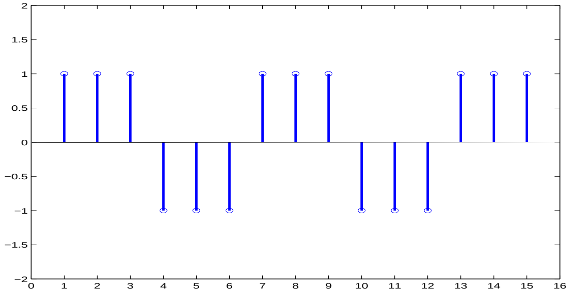
$$\bar{R}_n := \begin{bmatrix} R_u(0) & R_u(1) & \cdots & R_u(n-1) \\ R_u(1) & R_u(0) & \cdots & R_u(n-2) \\ \vdots & \ddots & \ddots & \vdots \\ R_u(n-1) & \cdots & R_u(1) & R_u(0) \end{bmatrix}$$

with $R_u(\tau) := \bar{\mathbb{E}}\{u(t)u(t-\tau)\}$, is non-singular.

Example:

A white noise process ($R_u(\tau) = \delta(\tau)$) is persistently exciting of **any** order. In that case: $\bar{R}_n = I_n$

Example block signal



$$R_u(0) = 1 \quad R_u(1) = \frac{1}{3} \quad R_u(2) = -\frac{1}{3} \quad R_u(3) = -1$$

$$R_u(4) = -\frac{1}{3} \quad R_u(5) = \frac{1}{3} \quad R_u(6) = 1 \quad \text{etcetera}$$

$$\bar{R}_4 = \begin{bmatrix} 1 & \frac{1}{3} & -\frac{1}{3} & -1 \\ \frac{1}{3} & 1 & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & 1 & \frac{1}{3} \\ -1 & -\frac{1}{3} & \frac{1}{3} & 1 \end{bmatrix}$$

$\bar{R}_3$ is regular, $\bar{R}_4$ is singular
so u is p.e. of order 3

A step signal:

$$
\begin{aligned}
u(t) &= 0 \quad t < 0 \\
&= c \quad t \geq 0
\end{aligned}
$$

is actually a transient signal, and therefore does not fit well into the "persistent signals" framework.
Although a step response reveals all dynamics of an LTI system, a step signal is p.e. of order 1, only having a spectral contribution at $\omega = 0$.
So it only persistently excites frequency 0.

System in the model set

Data-generating system $\mathcal{S}$: $[G_0, H_0]$
Model set: $\mathcal{M}$: $\{[G(q,\theta), H(q,\theta)], \theta \in \Theta \subset \mathbb{R}^d\}$.

$$\mathcal{S} \in \mathcal{M}$$

denotes that the data generating system can exactly be represented within $\mathcal{M}$, i.e. $\exists \theta_0 \in \Theta$ such that

$$\begin{aligned} G(q,\theta_0) &= G_0(q) \\ H(q,\theta_0) &= H_0(q) \end{aligned}$$

The notion $G_0 \in \mathcal{G}$ with $\mathcal{G} = \{G(q,\theta), \theta \in \Theta \subset \mathbb{R}^d\}$ denotes that only G_0 can exactly be represented, i.e.

$$G(q,\theta_0) = G_0(q)$$

Example

ARMAX system:

$$G_0 = \frac{b_0 q^{-1}}{1 + a_0 q^{-1}}; \qquad H_0 = \frac{1 + c_0 q^{-1}}{1 + a_0 q^{-1}}$$

ARX model:

$$G(q, \theta) = \frac{b q^{-1}}{1 + a q^{-1}}; \qquad H(q, \theta) = \frac{1}{1 + a q^{-1}}$$

Observation:

$$\mathcal{S} \notin \mathcal{M}; \qquad G_0 \in \mathcal{G}$$

Consistency result for $\mathcal{S} \in \mathcal{M}$

Consistency for $\mathcal{S} \in \mathcal{M}$

Let data be generated by a system $\mathcal{S} = (G_0, H_0)$, and consider a model set $\mathcal{M}$ with $G(q, \theta)$ parametrized according to

$$G(q,\theta) = q^{-n_k} \cdot \frac{b_0 + b_1 q^{-1} + \cdots + b_{n_b-1} q^{-n_b+1}}{1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}}.$$

If $\mathcal{S} \in \mathcal{M}$ and u is persistently exciting of order $\geq n_f + n_b$ then

$$\begin{aligned} G(e^{i\omega}, \theta^*) &= G_0(e^{i\omega}) \\ H(e^{i\omega}, \theta^*) &= H_0(e^{i\omega}) \qquad -\pi \leq \omega \leq \pi \end{aligned}$$

Note that n_b and n_f are number of parameters in numerator/denominator of $G(q, \theta)$.

Consistency result for $G_0 \in \mathcal{G}$

Consistency result for $G_0 \in \mathcal{G}$

Consider the same situation as on the previous slide.
If

- $G_0 \in \mathcal{G}$, and
- G and H are independently parametrized in $\mathcal{M}$, and
- u is persistently exciting of order $\geq n_f + n_b$

then

$$G(e^{i\omega}, \theta^*) \;=\; G_0(e^{i\omega}) \qquad -\pi \leq \omega \leq \pi$$

G_0 can be estimated consistently, irrespective of the question whether H_0 can be modeled exactly.

Independent parametrizations

	Model structure	$G(q,\theta)$	$H(q,\theta)$
	ARX	$\frac{B(q^{-1},\theta)}{A(q^{-1},\theta)}$	$\frac{1}{A(q^{-1},\theta)}$
	ARMAX	$\frac{B(q^{-1},\theta)}{A(q^{-1},\theta)}$	$\frac{C(q^{-1},\theta)}{A(q^{-1},\theta)}$
$\Longrightarrow$	OE - Output Error	$\frac{B(q^{-1},\theta)}{F(q^{-1},\theta)}$	1
$\Longrightarrow$	FIR	$B(q^{-1},\theta)$	1
$\Longrightarrow$	BJ - Box-Jenkins	$\frac{B(q^{-1},\theta)}{F(q^{-1},\theta)}$	$\frac{C(q^{-1},\theta)}{D(q^{-1},\theta)}$

Justification of consistency results:

From slide 4-31:

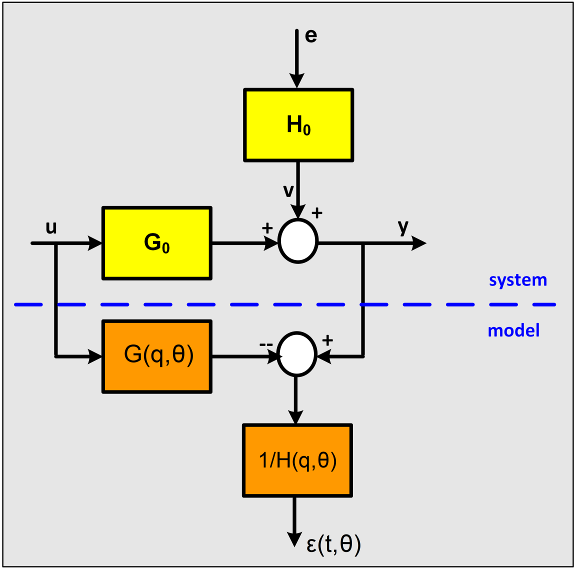
$$\varepsilon(t,\theta) = \frac{G_0(q) - G(q,\theta)}{H(q,\theta)} u(t) + \frac{H_0(q)}{H(q,\theta)} e(t)$$

- Power of first term is 0 $\rightarrow G(q,\theta) = G_0(q)$ provided that u is sufficiently exciting, and minimization of the two terms can be done independently;
- If $G(q,\theta)$ and $H(q,\theta)$ share parameters the two terms become dependent.
  Still the minimum of $\bar{V}(\theta)$ can be achieved by $G(q,\theta) = G_0(q)$ and $H(q,\theta) = H_0(q)$.

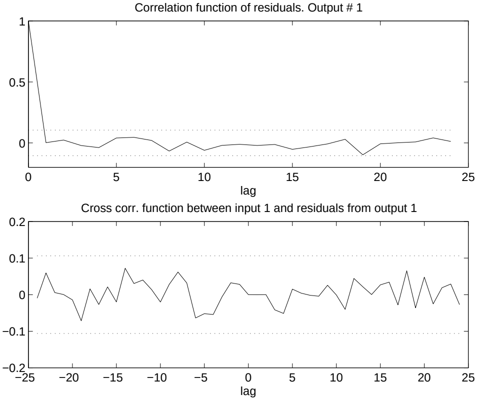
Validation for consistency

Can we determine whether a consistent estimation has been made?

- If G_0 and H_0 has been estimated consistently then $\varepsilon(t) = e(t)$ should be a realization of a white noise process $\rightarrow R_\varepsilon(\tau) = 0$ for $\tau \neq 0$.
- If G_0 has been estimated consistently then $R_{\varepsilon u}(\tau) = 0$ for all τ.

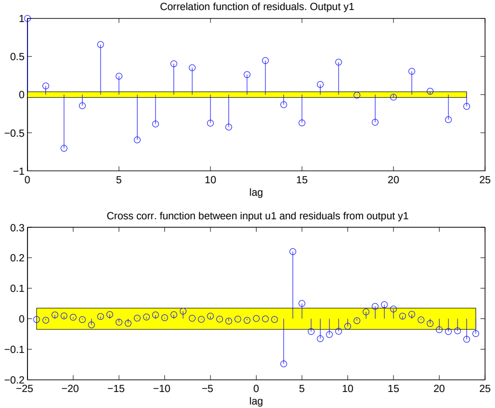

e
H_0
v
+
u
G_0
+
y
system
model
$G(q,\theta)$
--
+
$1/H(q,\theta)$
$\varepsilon(t,\theta)$

Through MATLAB-function RESID:



Confidence intervals of 99%.

On the basis of $N = 5000$ input-output data with white noise input select: $\mathcal{M} = OE(n_b = 2, n_f = 2, n_k = 3)$
and perform the residual test on the estimated model.

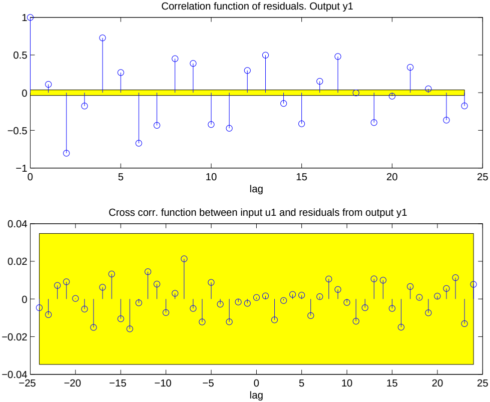


Both $G(\hat{\theta}_N)$ and $H(\hat{\theta}_N)$ are invalidated.

Increase the order of $G(q,\theta)$

$$\mathcal{M} = OE(n_b = 3, n_f = 3, n_k = 3)$$

and perform the residual test on the estimated model.

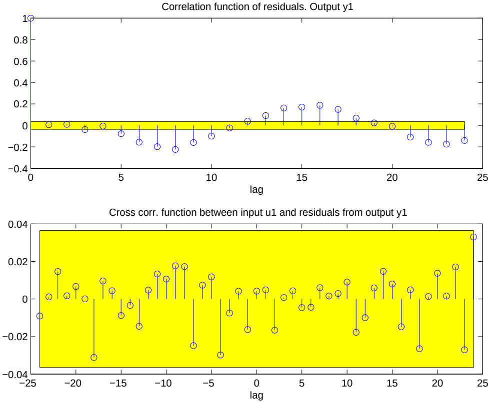


$G(\hat{\theta}_N)$ now is validated; $H(\hat{\theta}_N)$ is invalidated.

Extend the model structure with a noise model, BJ structure:

$$\mathcal{M} = BJ(n_b = 3, n_c = 3, n_d = 3, n_f = 3, n_k = 3)$$

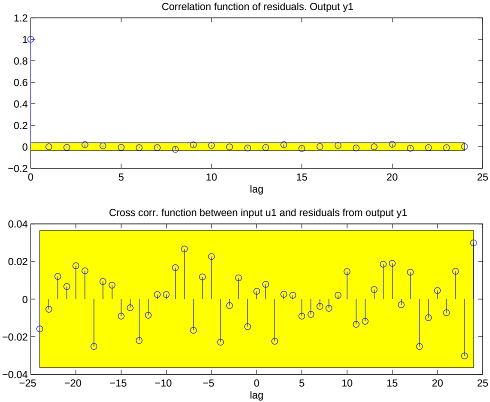
and validate:



$H(\hat{\theta}_N)$ is still invalidated.

Extend the order of the noise model to 4:

$$\mathcal{M} = BJ(n_b = 3, n_c = 4, n_d = 4, n_f = 3, n_k = 3)$$



Both $G(\hat{\theta}_N)$ and $H(\hat{\theta}_N)$ are now validated.
Note: the used $\mathcal{S}$ was indeed BJ(3,4,4,3,3) !!

What is the added value of estimating/validating H also?

Smaller variance of $G(q, \hat{\theta}_N)$ (will be addressed in Lecture 7).

Summary

- For $N \to \infty$ the estimated parameter $\hat{\theta}_N$ converges (w.p.1) to the asymptotic estimate $\theta^* = \arg\min_\theta \bar{\mathbb{E}}\varepsilon(t,\theta)^2$.
- The concepts “system in the model set $\mathcal{S} \in \mathcal{M}$ and $G_0 \in \mathcal{G}$ determine whether consistency can be achieved for (G, H) or for G respectively.
- If u is persistently exciting of a sufficiently high order then (G, H) is estimated consistently when $\mathcal{S} \in \mathcal{M}$.
- If u is persistently of a sufficiently high order and $G(\theta)$ and $H(\theta)$ are parametrized independently, then G is estimated consistently when $G_0 \in \mathcal{G}$.
- Residual tests can be used to “verify” consistency