System Identification
Lecture 8
Structure Selection and Model Validation

Paul Van den Hof

Control Systems Group
Department of Electrical Engineering
Eindhoven University of Technology

Contents

Structure Selection and Model Validation

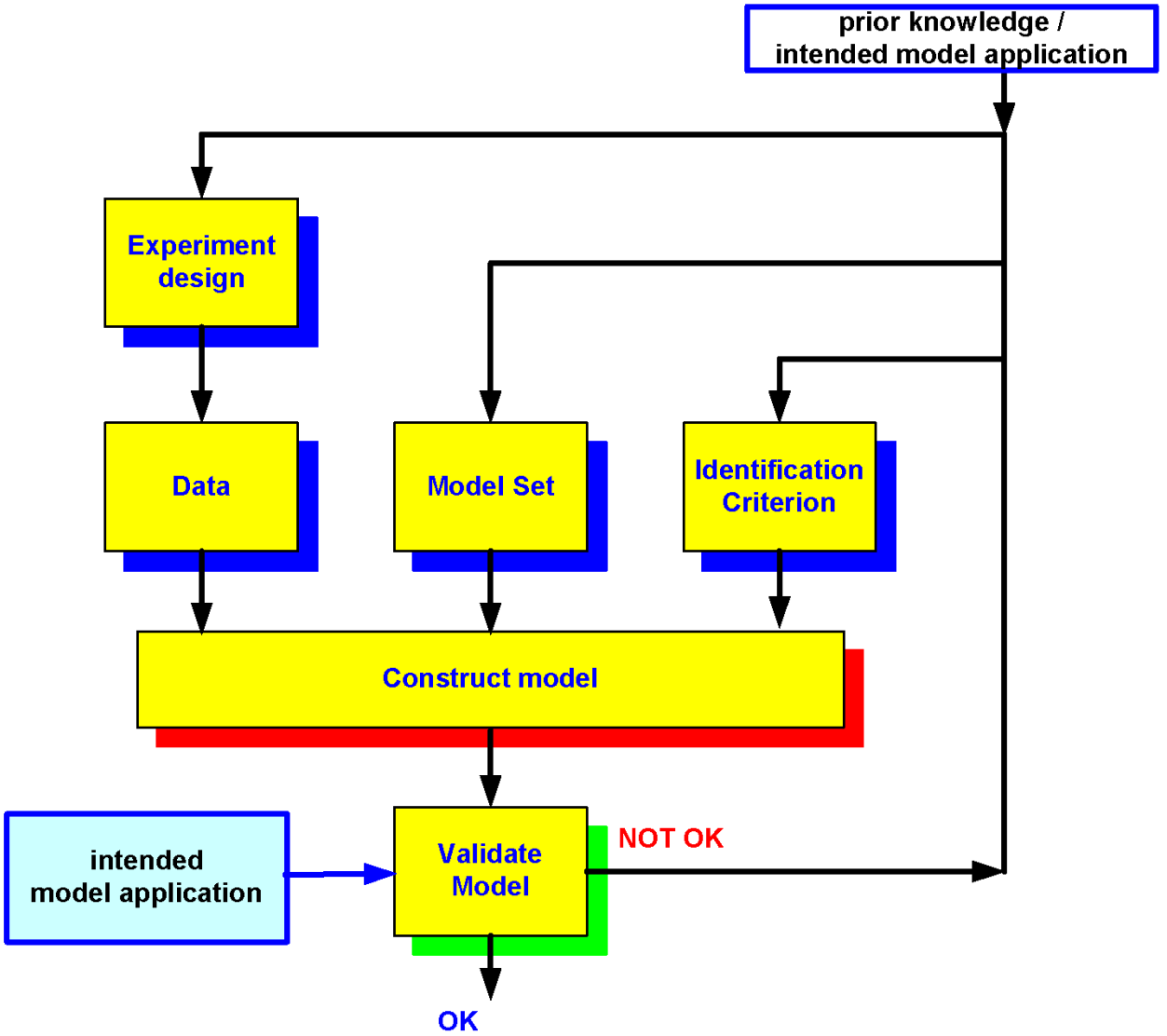
- How do we select an appropriate model structure and the orders of the different polynomials?
- How can we validate our identified models?

Model structure selection

Arguments for choice of model structure (ARX, OE, BJ)

- Algorithmic:
  Analytical solution (ARX, FIR) versus gradient-type algorithms (rest)
- Prior knowledge of physical location of noise influence
- Aspects of approximate modeling: weighting functions
- Decoupling of estimation of G_0 and H_0:
  independent parametrization (OE,FIR,BJ) versus dependent (rest)
- Minimum variance ($\mathcal{S} \in \mathcal{M}$)

In general however: Structure selection after model validation



Adapt the model set iteratively until a satisfying result is obtained

Global considerations

Choice of model complexity (order):

Large $\mathcal{M}$

- Good fit on data (small residual)
- Model is dependent on particular noise realization
- Large variance (in case of too many parameters in G and/or H)

Small $\mathcal{M}$

- Bad fit on data (large residual)
- Large bias

A priori considerations

- Relation between
  N number of data, and
  N_θ number of parameters to be estimated.
  Rule of thumb:

$$N > 10 \cdot N_\theta$$

  General: dependent on S/N-ratio

Comparing models (a posteriori)- Cross-validation

For a given structure, compare

$$V_N(\hat{\theta}_N, Z^N)$$

for $\hat{\theta}_N$ estimated for several model orders.
$\Rightarrow$ Large order leads to small V_N.

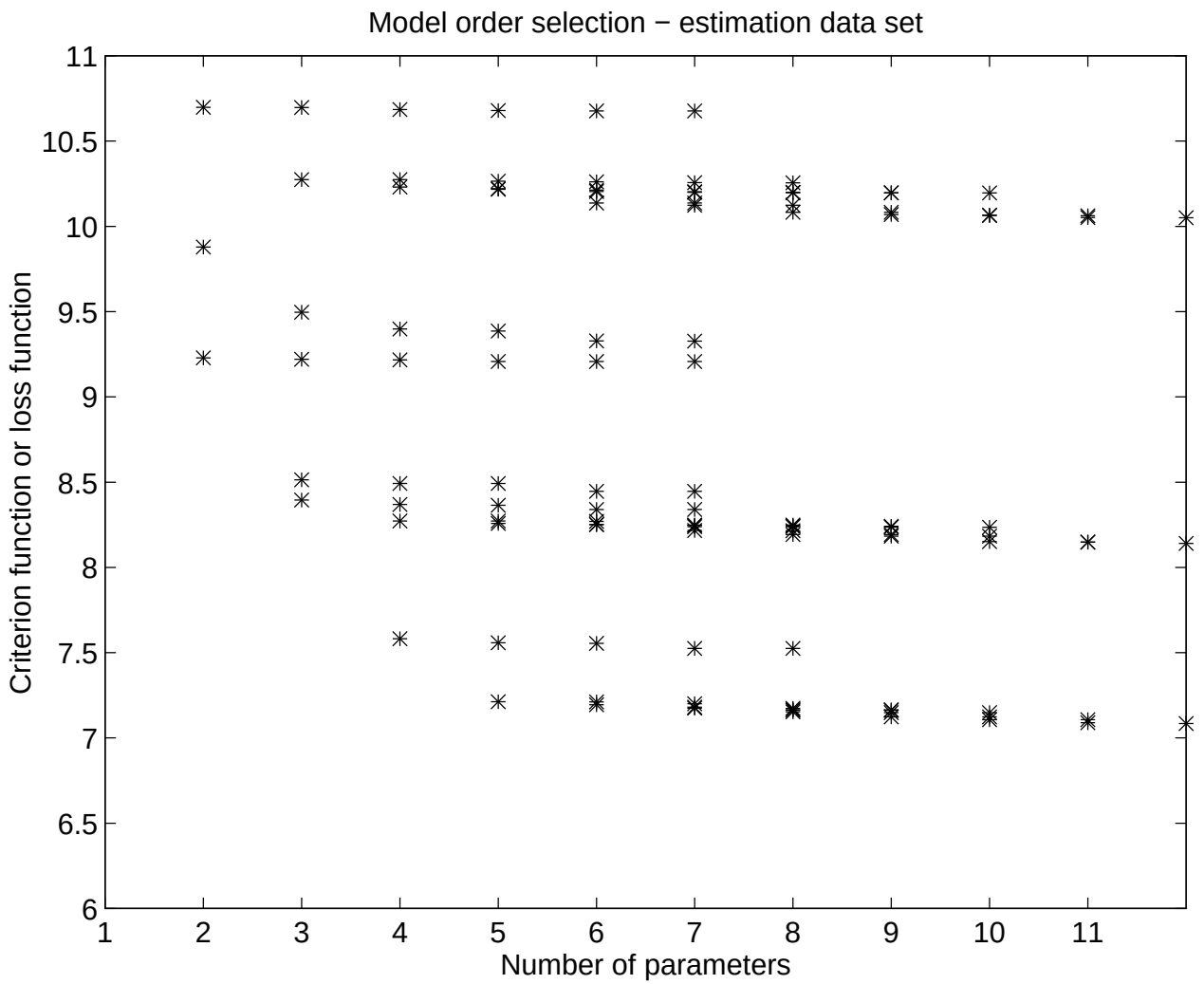
Avoiding of "overfit": Cross validation $Z^N = Z^{(1)}\ Z^{(2)}$
Estimate model on data $Z^{(1)}$:

$$\hat{\theta}_N^{(1)} = \arg\min_{\theta\in\Theta} V_N(\theta, Z^{(1)})$$

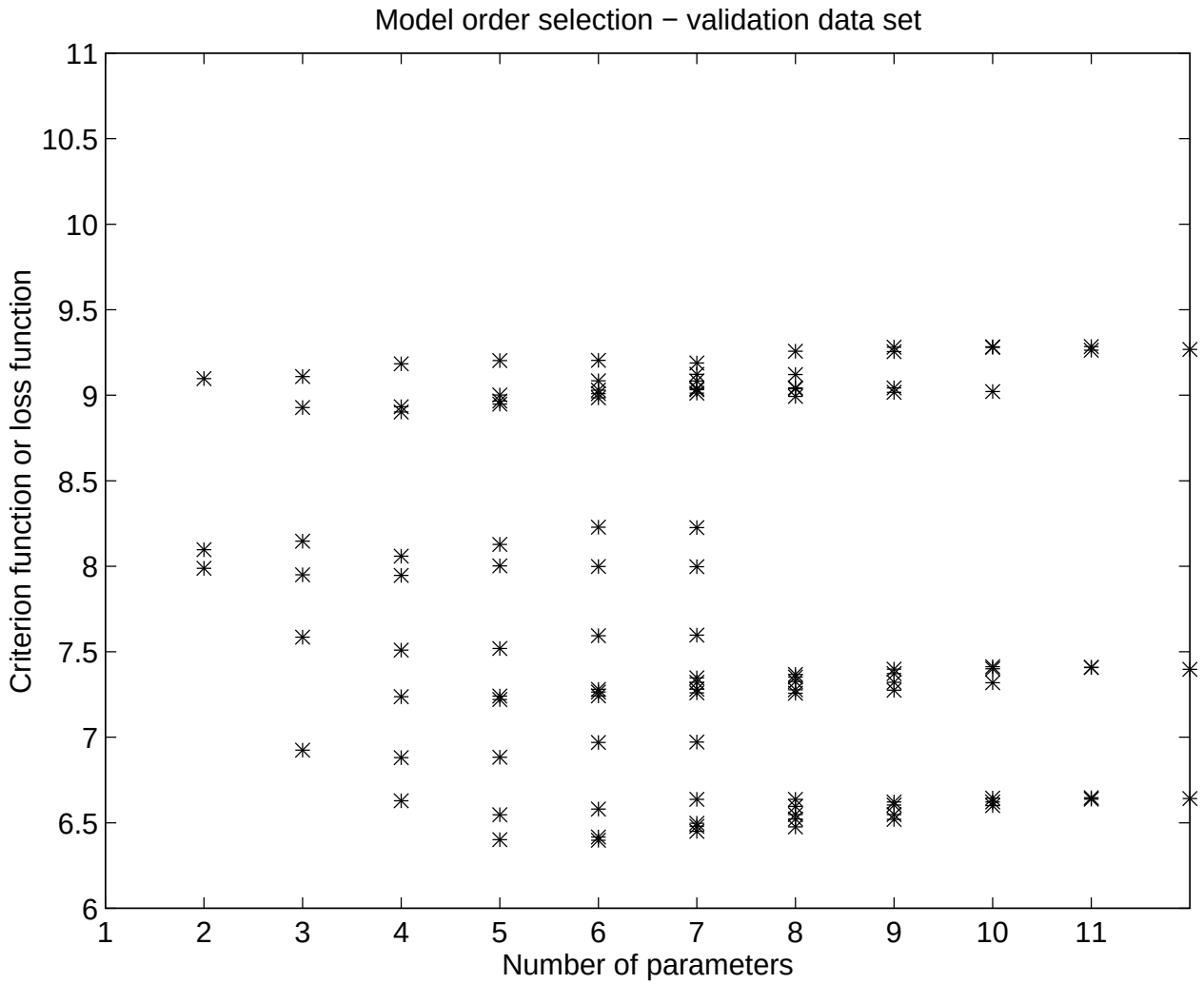
Evaluate criterion on data $Z^{(2)}$:

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

Loss function calculated on estimation data: $\min_{\theta \in \Theta} V_N(\theta, Z^{(1)})$



Loss function calculated on validation data: $V_N(\hat{\theta}_N^{(1)}, Z^{(2)})$



Model order selection on the basis of estimation data is tricky since a higher order always leads to a better fit.

Using validation data can lead to better order estimates, as now a penalty in the cost function appears in the case of overfit.

Note that for the validation data case, the "overfit" argument only works in the case of $\mathcal{S} \in \mathcal{M}$, i.e. all "model" information is already extracted from the data.

Model validation

Major question:

Can we validate/invalidate the identified model after an identification step?

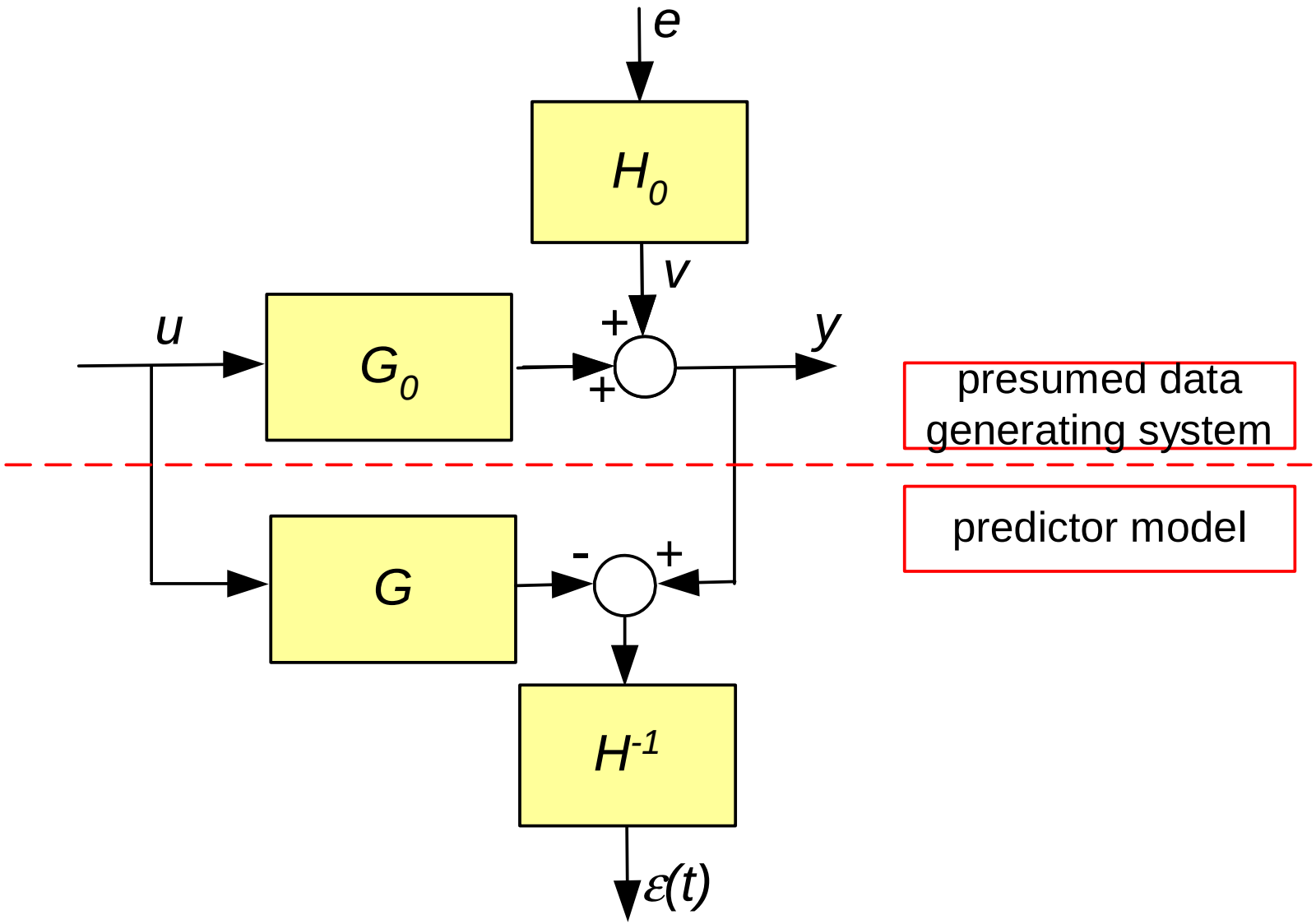
Options:

- $G(q, \hat{\theta}_N)$ in accordance with previous results?
  (non-parametric model; transient analysis; a priori info)
- Model reduction; pole-zero cancellations?
- Confidence intervals of estimated parameters
- Model simulation:

$$y_{sim}(t) = G(q, \hat{\theta}_N)u(t)$$

  Simulation error: $e_{sim}(t) = y(t) - y_{sim}(t)$
  In best case $e_{sim}(t) = v(t)$ (and not 0) Here also a risk of "overfit".

- Residual-tests
- Confidence intervals of $G(e^{i\omega}, \hat{\theta}_N)$


e
H_0
v
u
G_0
+
+
y
presumed data generating system
predictor model
G
-
+
H^{-1}
$\varepsilon(t)$

Note that the *simulation error* is different from the *prediction error*, except in the situation of an OE model structure

Residual-tests

$$\hat{R}_{\varepsilon}^{N}(\tau) := \frac{1}{N}\sum_{t=1}^{N-\tau}\varepsilon(t+\tau)\varepsilon(t) \qquad \hat{R}_{\varepsilon u}^{N}(\tau) := \frac{1}{N}\sum_{t=1}^{N-\tau}\varepsilon(t+\tau)u(t)$$

Test of hypotheses:

(a) $\varepsilon(t,\hat{\theta}_N)$ is a realization of a white noise process
(consistency $\hat{G}$, $\hat{H}$)
Evaluate

$$\hat{R}_{\varepsilon}^{N}(\tau) \to \delta(\tau)$$

(b) $\varepsilon(t,\hat{\theta}_N)$ is uncorrelated with past input samples
(consistency $\hat{G}$)
Evaluate

$$\hat{R}_{\varepsilon u}^{N}(\tau) \to 0, \qquad \tau \geq 0$$

Confidence intervals:

$$\sqrt{N}\frac{\hat{R}_\varepsilon^N(\tau)}{\hat{R}_\varepsilon^N(0)} \in As\ \mathcal{N}(0,1)$$

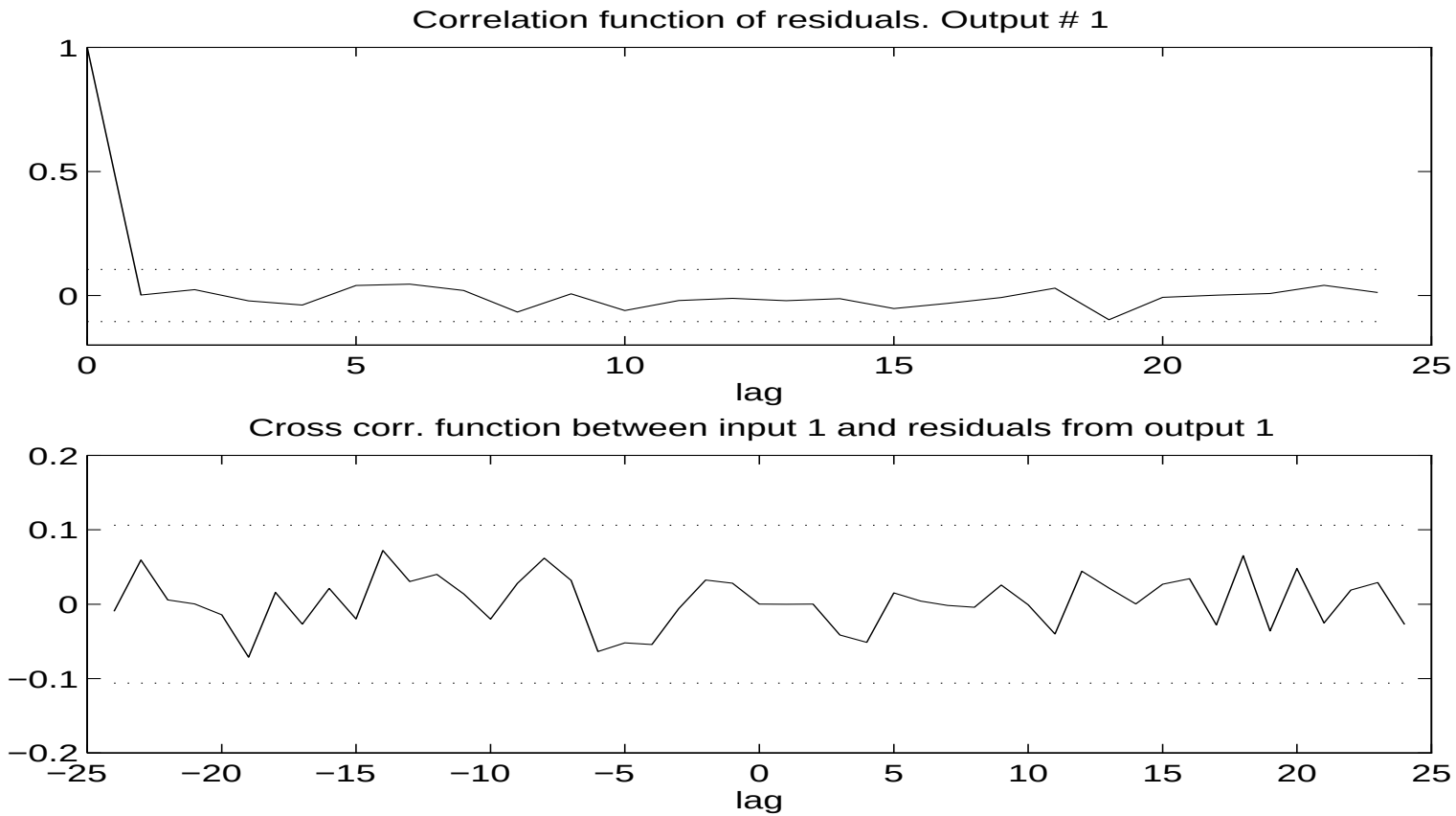
(valid for ε indeed being white noise)

$$\sqrt{N}\hat{R}_{\varepsilon u}^N(\tau) \in As\ \mathcal{N}(0,P)$$

$$P = \sum_{k=-\infty}^{\infty} R_\varepsilon(k)R_u(k)$$

(expressions for P based on the assumption that ε and u are indeed uncorrelated).

Through MATLAB-function RESID:



Confidence intervals of 99% (tunable in toolbox).

The pointwise (over time) test should actually be replaced by a vector test on the full vector

$$\begin{bmatrix} \hat{R}^N_{\varepsilon u}(0) & \hat{R}^N_{\varepsilon u}(1) & \cdots & \hat{R}^N_{\varepsilon u}(n_\tau) \end{bmatrix}^T$$

taking account of the correlations between the several terms.

S. Douma, X. Bombois and P.M.J. Van den Hof (2008). Validity of the standard cross-correlation test for model structure validation. *Automatica, Vol. 44*, no. 5, pp. 1285-1294, 2008.

Akaike's Information Criterion (AIC) (tentative)

Using cross-validation for comparing model sets

When distinguishing between estimation and validation data sets:

$\mathbb{E}V_N(\hat{\theta}_N)$: expected cost function on estimation data set

$\mathbb{E}\bar{V}(\hat{\theta}_N)$: expected cost function over validation data

taken over r.v. $\hat{\theta}_N$.

Then (Theorem 16.1, Ljung, 1999):

$$\mathbb{E}\bar{V}(\hat{\theta}_N) \approx \mathbb{E}V_N(\hat{\theta}_N) + \frac{1}{N} tr\, \bar{V}''(\theta^*) P_\theta$$

With $P_\theta = 2\sigma_e^2 \cdot [\bar{V}''(\theta_0)]^{-1}$ and $\theta^* = \theta_0$ it follows that

$$\mathbb{E}\bar{V}(\hat{\theta}_N) \approx V_N(\hat{\theta}_N) + \sigma_e^2 \frac{2n_\theta}{N} \quad \text{with } n_\theta = dim(\theta).$$

$$\mathbb{E}\bar{V}(\hat{\theta}_N) \approx V_N(\hat{\theta}_N) + \sigma_e^2 \frac{2n_\theta}{N} \quad \text{with } n_\theta = dim(\theta).$$

So, rather then minimizing $V_N(\theta)$ one should aim at minimizing

$$V_N(\theta) + \sigma_e^2 \frac{2n_\theta}{N}$$

i.e. extra penalty for "larger" model sets.

With σ_e^2 replaced by the estimate

$$\hat{\sigma}_e^2 = \frac{V_N(\hat{\theta}_N)}{1 - n_\theta/N}$$

the result is known as Akaike's Final Prediction Error criterion.

When applying the same reasoning to ML estimation with a Gaussian pdf, the resulting penalized cost function becomes

$$\log\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\theta)\right]+\frac{2n_\theta}{N}$$

which is known as Akaike's Information Criterion (AIC).

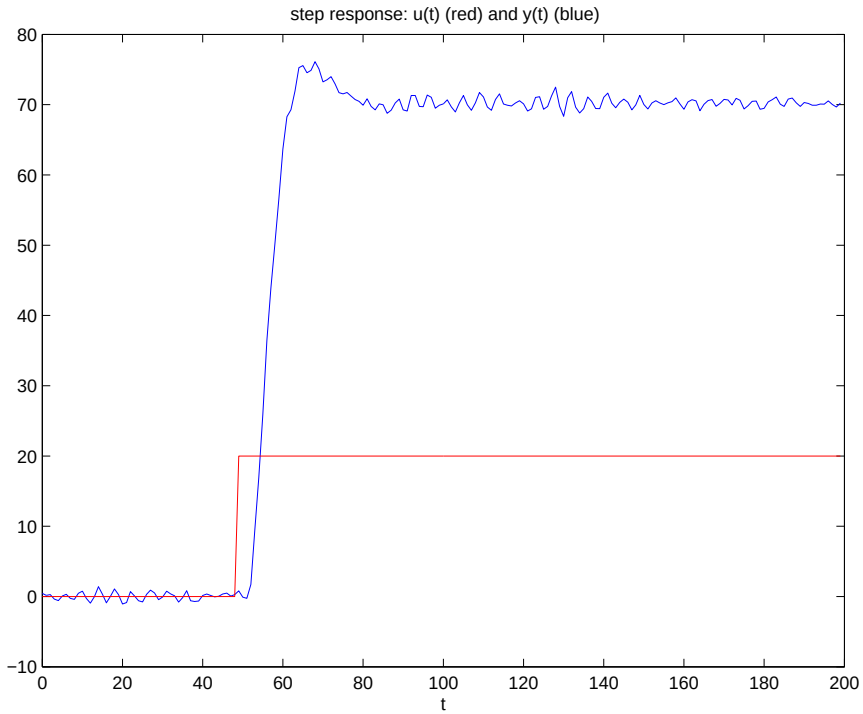
For further details see Ljung (1999), section 16.4.

Example of a model structure validation procedure
(repeated from Lecture 5)

Let us consider an unknown true system $\mathcal{S}$
We would like to determine a model set $\mathcal{M}$ that leads to a validated model.

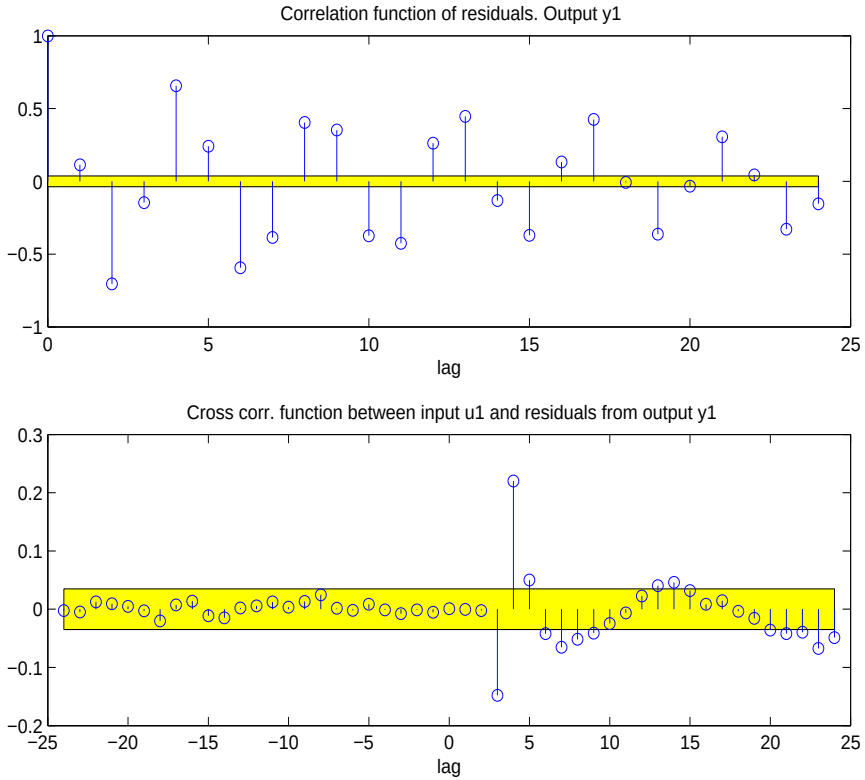
First analysis of the system

Let us apply a step input signal $u(t)$ to $\mathcal{S}$ and observe $y(t)$



"Conclusion": G_0 has a limited order and a delay of $n_k = 3$

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.

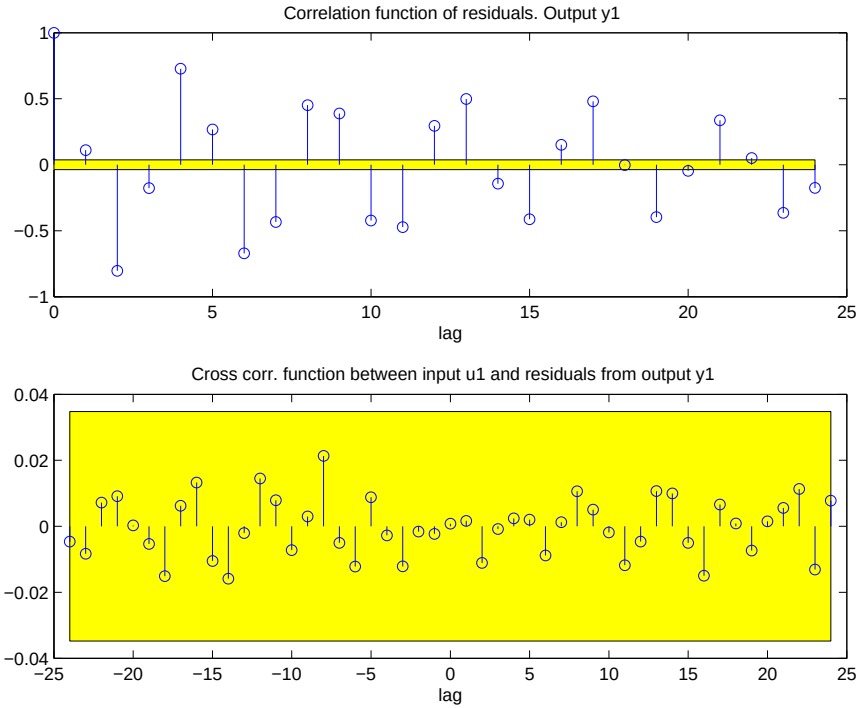


Both $G(\hat{\theta}_N)$ and $H(\hat{\theta}_N)$ are invalidated.

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.

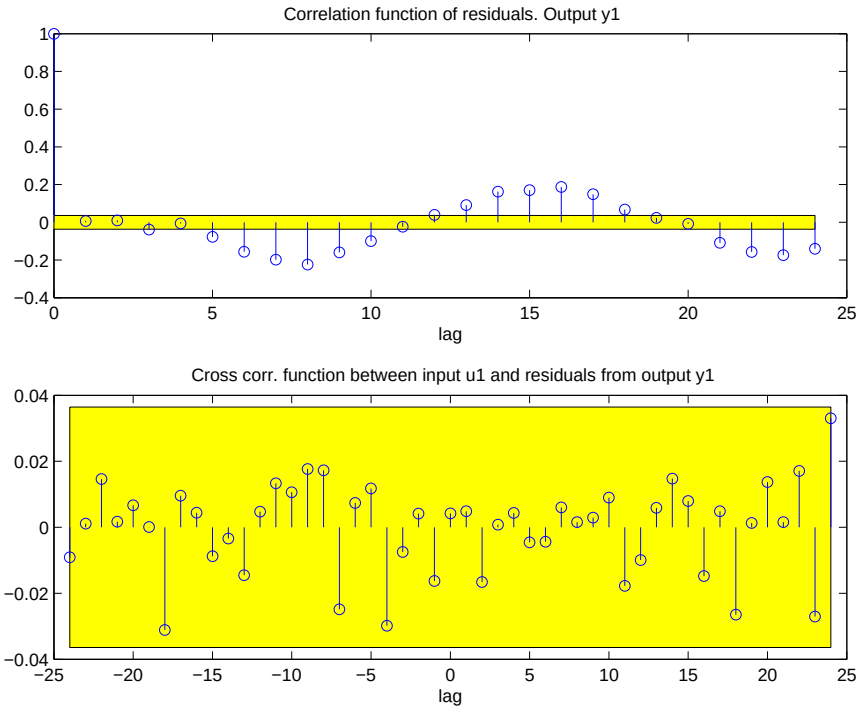


$G(\hat{\theta}_N)$ now is validated; $H(\hat{\theta}_N)$ is invalidated.

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

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

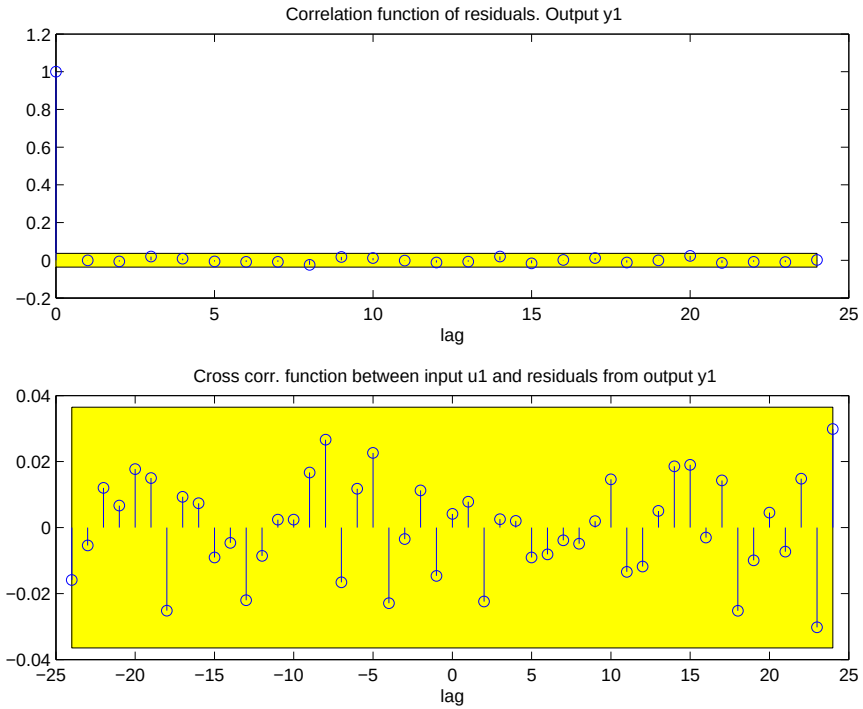
and validate:



$H(\hat{\theta}_N)$ 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) !!

A "standard" identification problem

- After preparatory experiments, estimate a non-parametric model on the basis of fastly sampled data
- Choose the correct sample interval for parametric models
- Estimate a (high order) ARX/OE model aiming first at a correct identification of G_0; evaluate with cross validation test;
- When using ARX, be aware of pole/zero cancellations in $G(\hat{\theta}_N)$
- Determine the order for a suitable OE model (for refining $\hat{G}$)
- When $\hat{G}$ is validated, extend the model with a noise part $H(q, \theta)$ through e.g. a Box-Jenkins structure, and validate the result through the autocorrelation test.

Validation through confidence intervals on $G(e^{i\omega}, \hat{\theta}_N)$

(tentative)

$$cov(G(e^{i\omega}, \hat{\theta}_N)) := \mathbb{E}\left(|G(e^{i\omega}, \hat{\theta}_N) - G(e^{i\omega}, \theta_0)|^2\right)$$

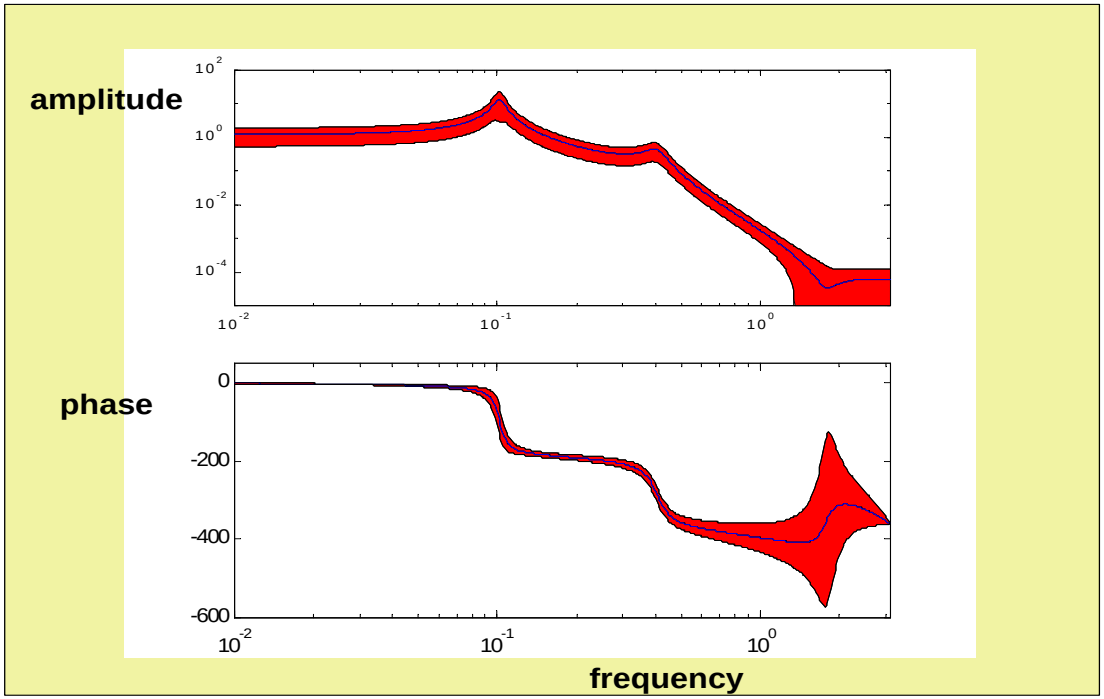
Consequently, at each frequency ω:
the modeling error $|G(e^{i\omega}, \theta_0) - G(e^{i\omega}, \hat{\theta}_N)|$ is very likely to be small
if
the standard deviation $\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ of $G(e^{i\omega}, \hat{\theta}_N)$ is small

More precisely, we have at each frequency ω that

$$|G(e^{i\omega}, \hat{\theta}_N) - G(e^{i\omega}, \theta_0)| < 2.45 \sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)} \quad w.p.\ 95\%$$

(generalization of the 2σ confidence intervals for complex-valued normal distribution)

$\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ is thus a measure of the modeling error and allows to deduce uncertainty bands around the frequency response of the identified model $G(q, \hat{\theta}_N)$



What is a small standard deviation $\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ (or a small modeling error)?

Highly dependent on the expected use of the model !!

For example, if we want to use the model for control, the modeling error (measured by $\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N))}$ has to be small up to the bandwidth of the closed loop.

A rule of thumb is $\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)} < 0.1|G(e^{i\omega}, \hat{\theta}_N)|$ in the bandwidth.

See the literature on "identification for robust control" to know how large $\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ may be.

What to do if the variance/uncertainty appears too large ?

If the variance $cov(G(e^{i\omega}, \hat{\theta}_N))$ appears too large, then we can not guarantee that $G(q, \hat{\theta}_N)$ is a close estimate of $G_0(q)$.

A new identification experiment has to be performed in order to obtain a better model with reduced variance.

Example

Let us consider the same flexible transmission system $\mathcal{S}$ (in the ARX form),
and a full order model structure $\mathcal{M}$, such that $\mathcal{S} \in \mathcal{M}$.

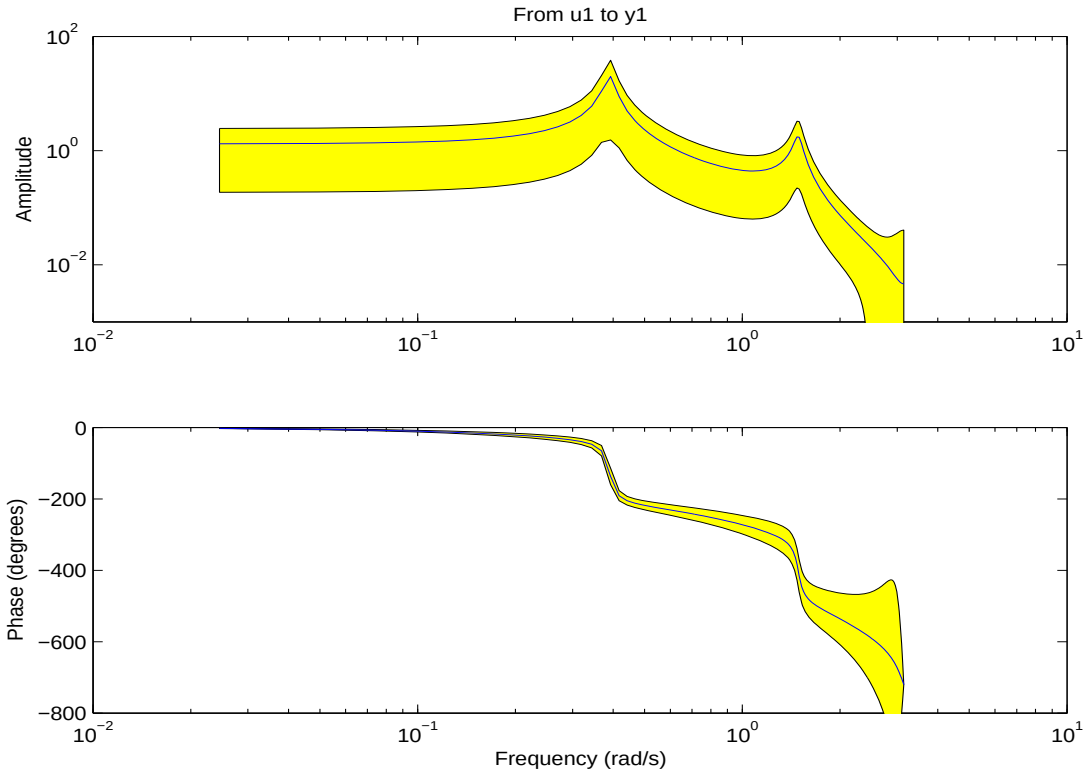
We intend to use $G(q, \hat{\theta}_N)$ for model-based control, where we need

$$\frac{\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}}{|G(e^{i\omega}, \hat{\theta}_N)|} < 0.1 \quad \forall \omega \in [0\ 1].$$

First identification experiment

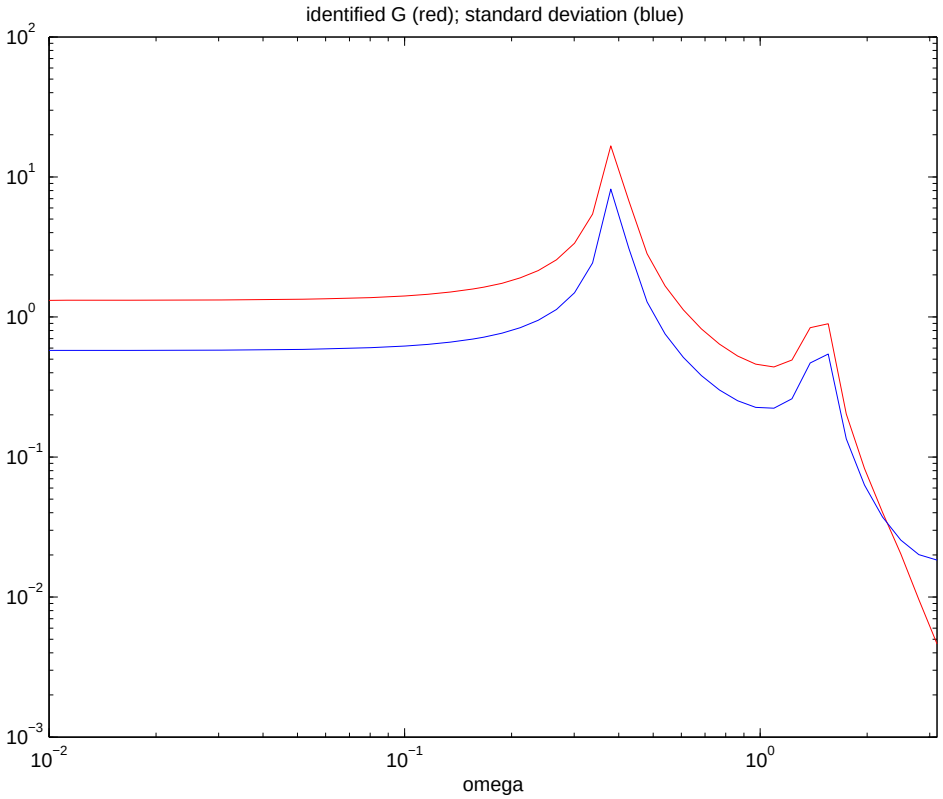
We apply a white noise input signal $u(t)$ of variance $\sigma_u^2 = 0.005$ to $\mathcal{S}$, collect $N = 2000$ IO data and identify a model $G(q, \hat{\theta}_N)$ in $\mathcal{M}$.

Using these data, we identify a model $G(e^{i\omega}, \hat{\theta}_N)$ (blue) and estimate its uncertainty with $\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ (yellow band):



The uncertainty seems quite large

To verify if the model is validated, we compare $\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ (blue) and $|G(e^{i\omega}, \hat{\theta}_N)|$ (red):



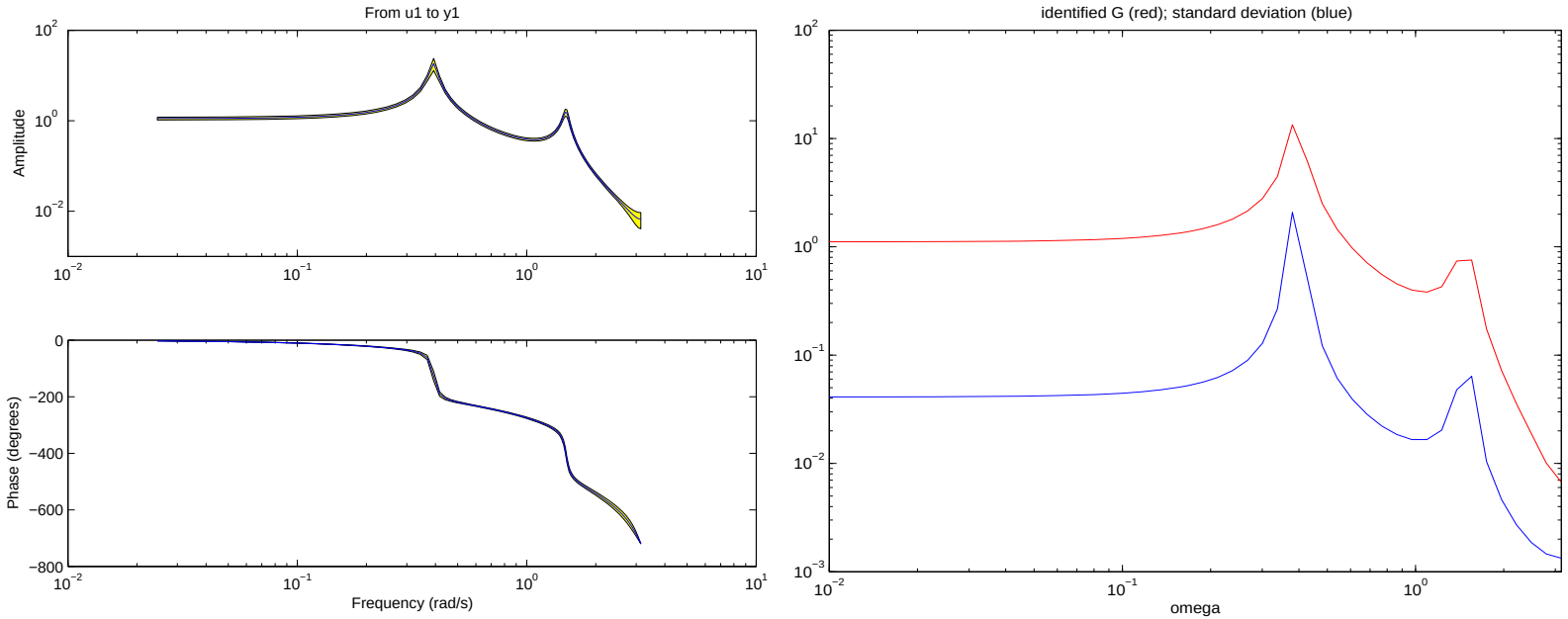
$\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ is too large !!!

Second identification experiment

We need to reduce the variance of the identified model.

For this purpose we increase the power of the white noise input $u(t)$ to $\sigma_u^2 = 1$,
and we collect $N = 2000$ IO data to identify a model $G(\hat{\theta}_N)$.

Validation of the identified model $G(e^{i\omega}, \hat{\theta}_N)$:



$\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ is better, but still too large at the 1st resonance peak.

Third identification experiment

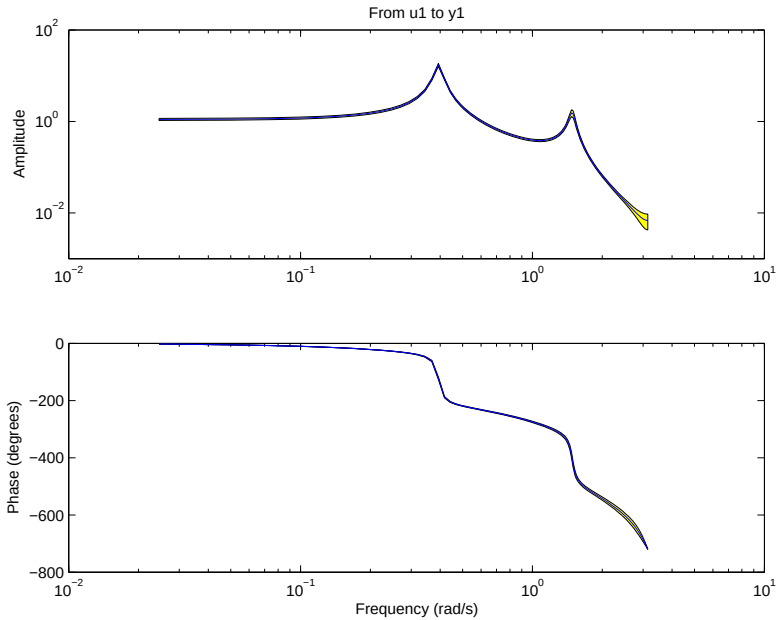
We need to reduce the variance of the identified model further around the 1st resonance peak.

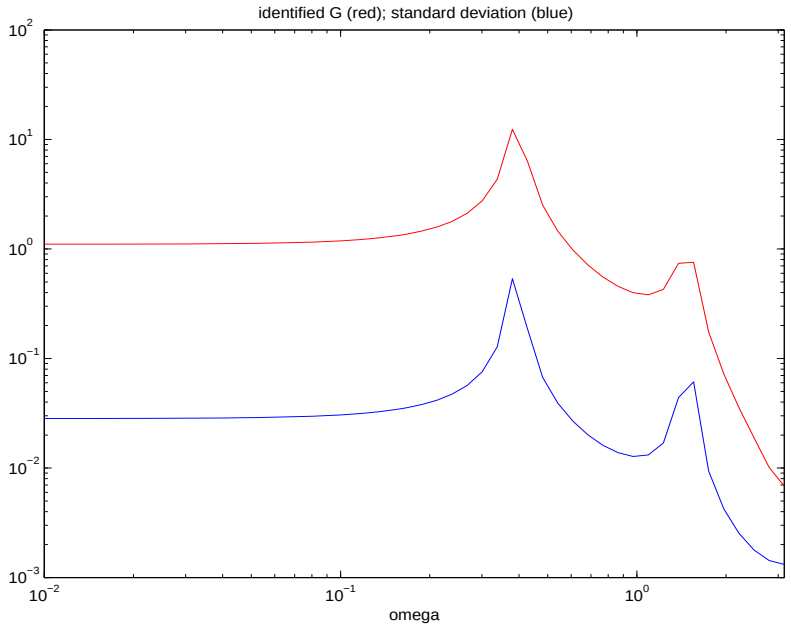
For this purpose we increase the power of $u(t)$ around this first peak, by choosing:

$$u(t) = \text{ white noise of the 2nd experiment } + sin(0.3t) + sin(0.4t)$$

We apply this input signal $u(t)$ to $\mathcal{S}$, collect $N = 2000$ IO data and identify a model $G(\hat{\theta}_N)$ in $\mathcal{M}$.

Validation of the identified model $G(e^{i\omega}, \hat{\theta}_N)$:





$\sqrt{cov(G(e^{i\omega}, \hat{\theta}_N)}$ is now OK for our control purpose!!!

In "classical" identification, estimation in low/fixed order model structures is standard.

Newer developments (see e.g. the course Machine learning for control) extend beyond these concepts towards high-order/nonparametric models with regularizing smoothness constraints.

Summary

- Selection of model structure based on several properties of consistency, approximation and computational aspects
- "Standard" procedure available to first focus on G and later on H
- Order selection typically through model comparisons (with validation data) or through the process of validation
- Several model validation methods have been presented, among which the residuals tests.
- The ultimate model validation is in the *use* of the model.