System Identification

sc4110

X. Bombois
P.M.J. Van den Hof

Material:

- Lecture notes sc4110 - January 2006
  available through: Blackboard or Nextprint

Lecture hours: (see schedule for details)

- Monday (15:45-17:30) in Room D (3Me)
- Thursday (15:45-17:30) in Room D (3Me)
- Friday (13:45-15:30) in Room A (API)

Part I: Introduction to system identification

System identification is about modeling

Notion of model common in many branches of science

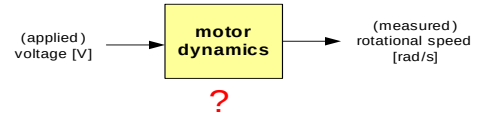
Within (systems and control) engineering:
models of dynamical systems for the purpose of

- system (re)design
- control design
- prediction
- simulation
- diagnosis / fault detection

System identification is about data-based modeling

data-based modeling ???

Data-based modeling of the DCSC DC motor

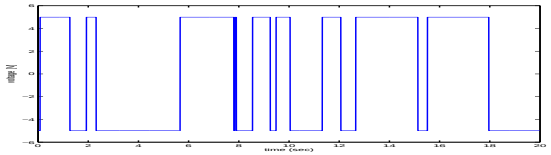


Determine a model of the dynamical relation existing between the voltage $u(t)$ driving the motor and the angular speed $y(t)$ of the rotor

How to proceed?

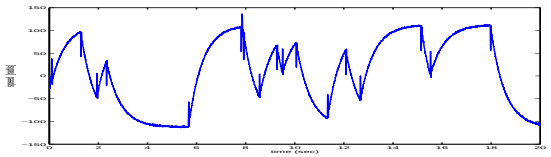
- Excite the system by applying the following sequence for the voltage $u(t)$ during 20 seconds

applied $u(t)$

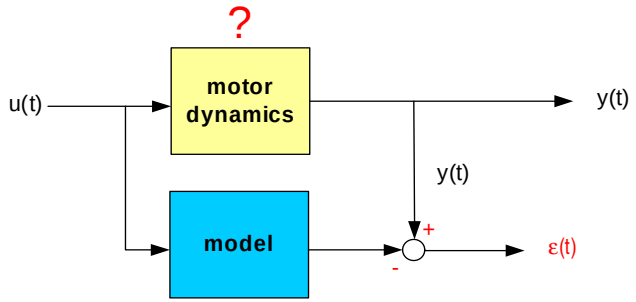


- Measure the induced rotational speed $y(t)$

measured $y(t)$



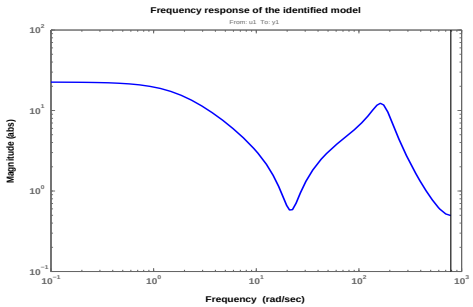
- Given a candidate model (i.e. a transfer function), we can use the available data to compute the signal $\varepsilon(t)$ featuring the modeling error



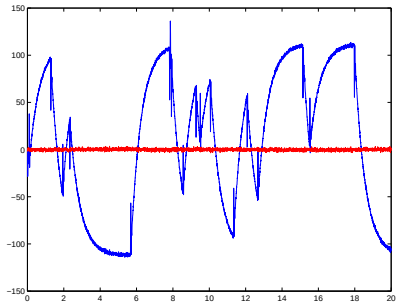
- determine that model minimizing the power of $\varepsilon(t)$ (often a filtered $\varepsilon(t)$; see later)

Identification result: a discrete-time transfer function (4^{th} order)

Frequency response



Measured $y(t)$ (blue)
vs. $\epsilon(t)$ (red)



$\varepsilon(t)$ contains not only the model inaccuracy, but also the noise acting on the system

Why is data-based modeling useful ?

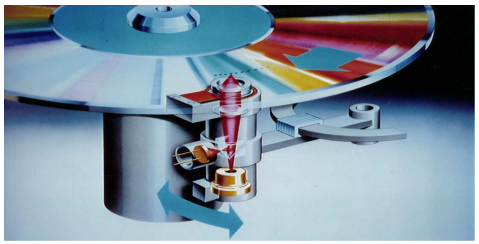
When thinking of modeling, we indeed generally think of first-principle modeling and not data-based modeling

first-principle modeling = modeling using the laws of physics (Newton, mass conservation..)

However, data-based modeling is often as important as first-principle modeling

Example 1: control of the pick-up mechanism of a CD-player

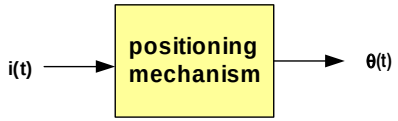
pick up mechanism: position the reading tool (laser) on the right track of the CD using a mechanical arm



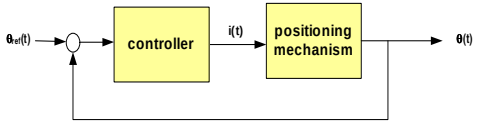
Arm is driven by the current $i(t)$ of a motor

Optical sensor to measure the laser position $\theta(t)$

Dynamical system



Objective: design a fast and precise position controller (required bandwidth $\approx 1000\ Hz$)



$\Longrightarrow$ model is needed

First-principle modeling

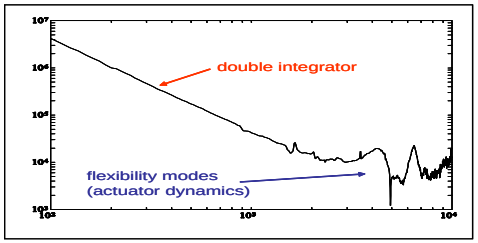
The model is designed based on the Newton law

Since the current induces a force, the relation between the current and the position is modeled by a double integrator

The controller designed with this physical model could not achieve the desired bandwidth without thrilling

Data-based modeling

An identification experiment was then performed and the following model identified



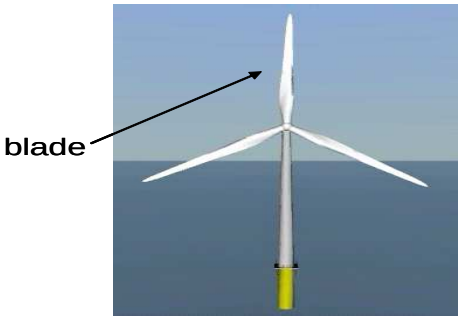
For a bandwidth of $\approx 1000\ Hz$, the mechanical modes can no longer be neglected and should be tackled by the controller

These flexible modes are quasi impossible to model with physical laws

Identified model $\Longrightarrow$ new controller design

Since all significant dynamics were now tackled, the controller based on the identified model showed satisfactory behaviour

Example 2: fatigue load reduction for new generation of wind turbines *J.W. Wingerden et al., Wind Energy, Wiley, 2008*

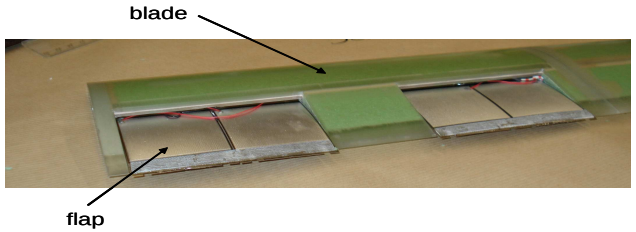


The blades of a wind turbine are subject to high vibration loads due to wind gust, periodic rotations, ...

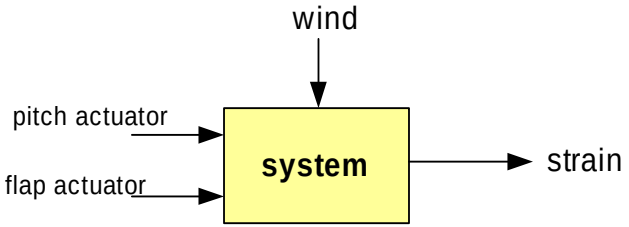
To enhance the life duration of wind turbines, these vibrations must be regulated

Two control loops to reduce the strain in the blade:

- pitch control: optimal orientation of the blades
- flap control: optimal orientation of flaps added to the blade structure



For control design, we need a model of the dynamics between the pitch and flap actuators and the strain in the blade:

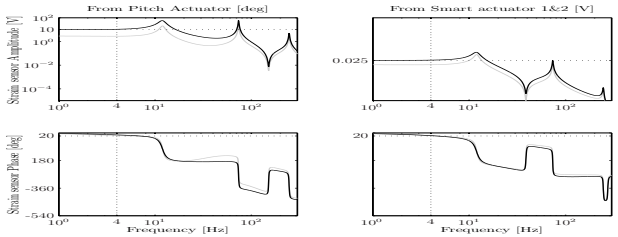


First-principle modeling

Model based on aerodynamic and mechanical laws

linear model (order $= 28$)

many physical parameters to determine $\Longrightarrow$ high uncertainty

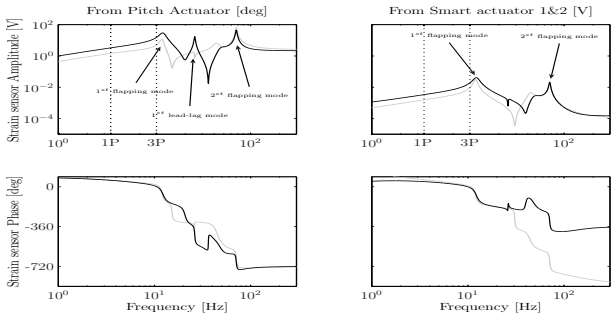


With this model, impossible to deduce a controller stabilizing the real-life system

Data-based modeling

We excite both inputs up to $100Hz$ (important band for control) and measure the corresponding strain

Based on these data, the following model is identified



Important differences between the two models

Behaviour in low frequencies (the physical model did not take into account the strain sensor dynamics)

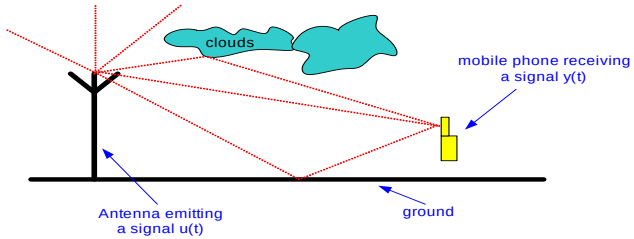
Extra resonance between 10 Hz and 100 Hz due to other vibration modes (unmodeled in the first-principle approach)

The identified model is simpler (order $= 10$) and less uncertain[a]

control design based on the identified model leads to a satisfactory reduction of the strain in the blade

[a] parameters of first-principle model have in fact been tuned with the identified model

Example 3: Signal equalization in mobile telephony



The received signal $y(t)$ is made up of several delayed versions of the emitted signal $u(t)$ + noise

$$y(t) = g_1\ u(t - n_1) + g_2\ u(t - n_2) + ... + \ noise$$

$\Longrightarrow$ distorted signal

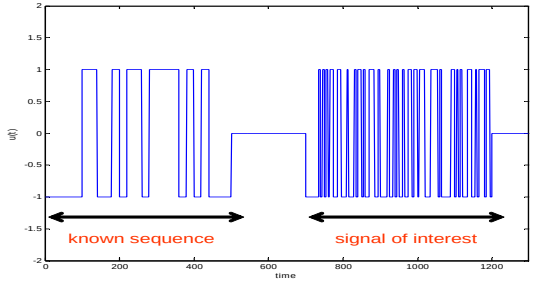
A model of the so-called channel is required to reconstruct $u(t)$ from the distorted $y(t)$

This model can not be determined in advance since the position of the mobile phone is mobile (by definition)

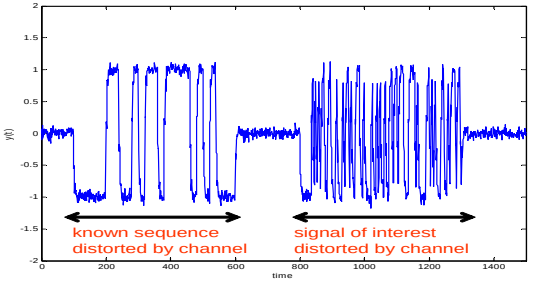
The model is identified at each received call

How to proceed?

When emitting $u(t)$, the signal of interest $u_{interest}(t)$ is preceded by a known sequence $u_{known}(t)$



Both the known sequence and the signal of interest are distorted by the channel



Denote by $y_{known}(t)$ and $y_{interest}(t)$ the received signals corresponding to $u_{known}(t)$ and $u_{interest}(t)$, respectively

Since u_{known} is a known sequence, the GSM software uses the data u_{known} and y_{known} to identify a model of the channel

This model can be then used to determine an appropriate filter to reconstruct $u_{interest}(t)$ from $y_{interest}(t)$

Summary: First-principle vs. Data-based modeling"

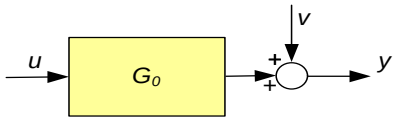
the two methodologies are often combined to increase confidence in the model

General disadvantages of first-principle modeling

- model contains many unknown (physical) parameters $\Longrightarrow$ high uncertainty (not quantifiable)
- model generally more complicate than with system identification
- missing actuator/sensor dynamics and phenomena can be forgotten
- sometimes impossible to determine (as in example 3, but also in the process industry)
- no disturbance model

System Identification: the players

to-be-identified system



$u(t)$ is the (discrete-time) input which can be freely chosen

$y(t)$ is the (discrete-time) output which can be measured and is made up of

- a contribution due to $u(t)$ i.e. $G_0 u(t)$
- a contribution independent of $u(t)$ i.e. the disturbance $v(t)$

the signal $v(t)$ is an unknown disturbance (noise, process disturbance, effects of non-measured inputs, ..)

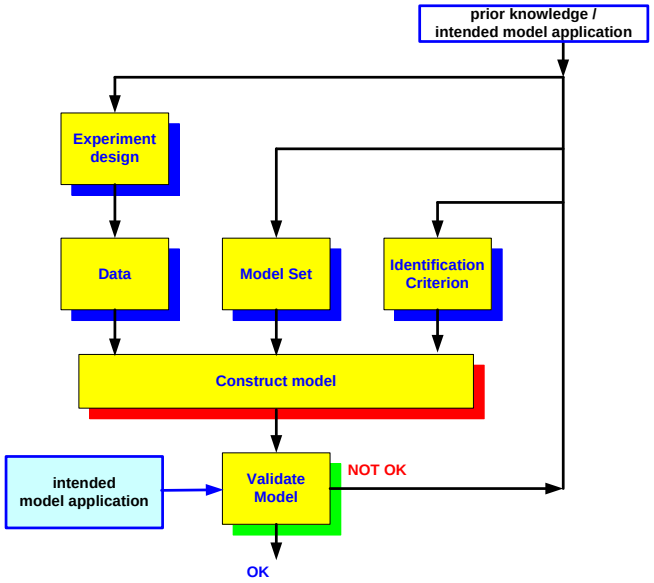
It can be best modeled via a (zero-mean) stochastic process. Indeed, $v(t)$ will never be the same if you repeat the experiment

The challenging nature of system identification is due to the presence of $v(t)$

If $v(t) = 0$, it is just an algebraic game to find the relation between $u(t)$ and $y(t)$

As result, an identification experiment (generally) delivers both a model of the transfer G_0 and of the disturbance $v(t)$

System identification procedure



Identification Criterion

Measure the "distance" between a data set $(u, y)_{t=1,\cdots N}$ and a particular model.

In this course, we will consider two criteria

- Prediction Error Identification (PEI) delivering a discrete-time transfer function as model of G_0
- Empirical Transfer Function Estimate (ETFE) delivering an estimate of the frequency response of G_0

Why those?

- PEI is the most used method in practice and the one delivering the most tools to validate a model
- ETFE is used to have a first idea of the system and facilitate the use of PEI

Other criteria: subspace identification, IV methods, ML methods, ...

Model set

Complexity of models (order, number of parameters) to be determined

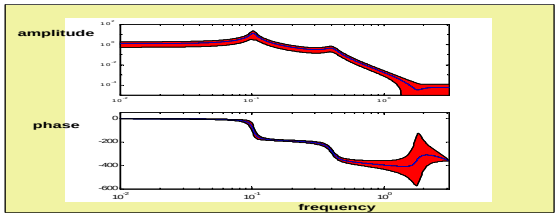
Experiment Design

- Choice of the type of excitation
  - sum of sinusoids (multisine)
  - realization of (filtered) white noise or alike
- Which frequency content?
- Which duration?

Experiment design is very important since it has a direct influence on the quality of the model

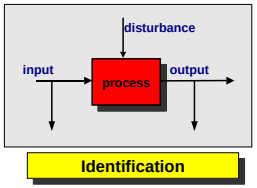
Model validation

- Comparing the actual output of the system with the output predicted by the model
- Determining the uncertainty of the system e.g. in the frequency domain

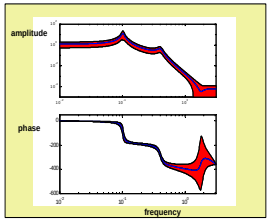


- ....

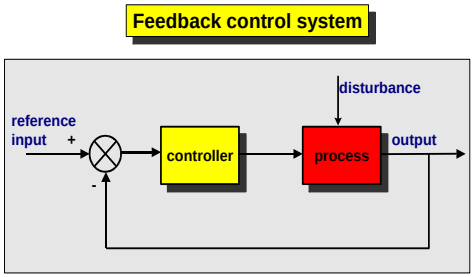
System identification for (robust) control



Data → Model



Model → Controller



History

- Basic principle (LS) from Gauss (1809)
- Development based on theories of
  - stochastic processes
  - statistics
- Strong growth in sixties and seventies
  Åström en Bohlin (1965), Åström en Eykhoff (1971)
- Brought to technological tools in nineties
  (Matlab Toolboxes for either time-domain of frequency domain),
  as well as to professional industrial control packages
  (Aspen, SMOC-PRO, IPCOS, Tai-Ji Control, AdaptX, ...).

Notions from estimation theory

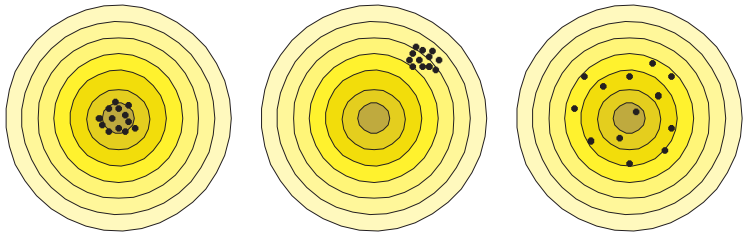
Estimator $\hat{\theta}_N$ of θ_0 based on N data points.

a. *Unbiased* (zuiver): $E\hat{\theta}_N = \theta_0$

b. *Consistent*. $\hat{\theta}_N$ is consistent if:

- $Pr[\lim_{N\to\infty} \hat{\theta}_N = \theta_0] = 1$
- $\hat{\theta}_N \to \theta_0$ with probability 1 voor $N \to \infty$.

c. *Variance*: $cov(\hat{\theta}_N) = E(\hat{\theta}_N - E\hat{\theta}_N)(\hat{\theta}_N - E\hat{\theta}_N)^T$.



Bull's eye represents θ_0;
left: unbiased estimate with small variance
middle: biased estimate with small variance
right: unbiased estimate with large variance

Part II: RECAP on discrete-time systems and signals

1. Introduction

Why are discrete-time systems and signals important in system identification?

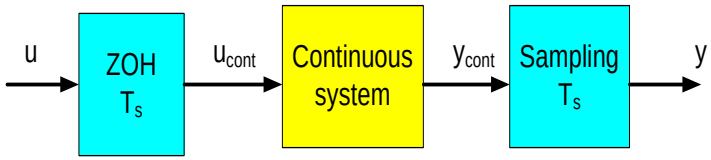
In system identification, we deal with measured signals $\Longrightarrow$ discrete-time signals

$\Longrightarrow$

the models/systems can be represented by discrete-time transfer functions

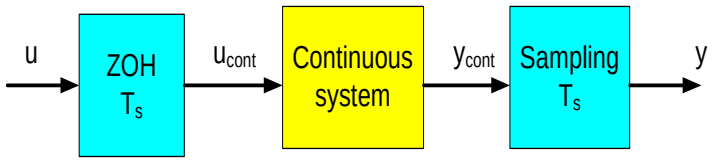
2. Discrete-time systems

Continuous-time vs. Discrete-time systems



The continuous output $y_{cont}(t_c)$ ($t_c \in \mathrm{R}$) of the system is sampled with sampled time T_s

This sampling delivers the discrete measurements $y(t)$ where $t = 0, 1, 2, \ldots$ i.e. $y(t) = y_{cont}(tT_s)$



The system is excited via the discrete sequence $u(t)$ $t = 0, 1, 2, \ldots$ generated by a PC

This discrete signal is made continuous by the Zero Order Hold (ZOH):

$$u_{cont}(t_c) = u(t) \quad \text{for } tT_s \leq t_c < (t+1)T_s$$

Illustration:

Continuous system: $G_0(s) = \frac{10}{s+10}$

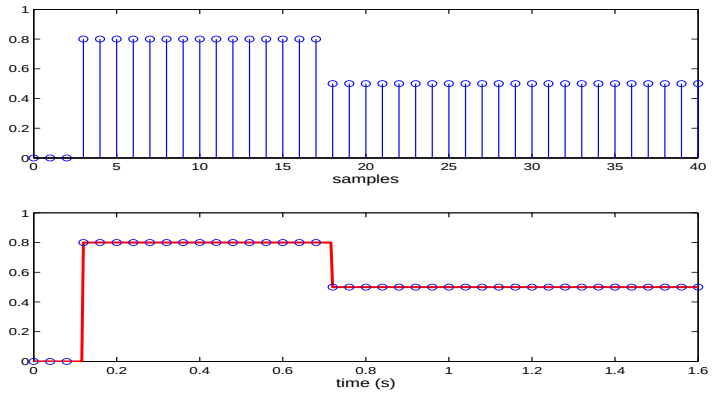
Sampling time: $T_s = 0.04\ s.$

The sequence $u(t)$ is made up of 41 samples i.e. $t = 0...40$

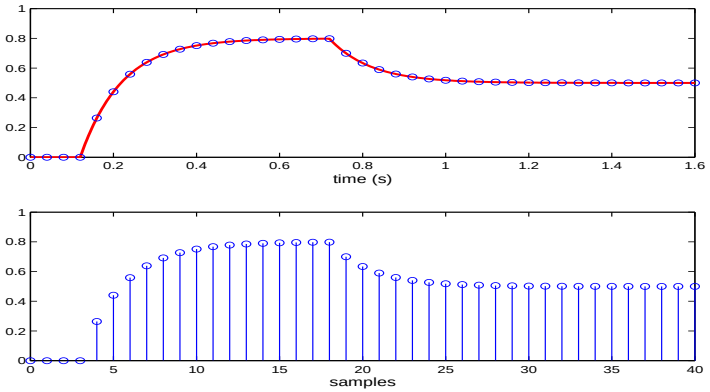
$$u(t) = \begin{cases} 0 & for\ 0 \leq t \leq 2 \\ 0.8 & for\ 3 \leq t \leq 17 \\ 0.5 & for\ 18 \leq t \leq 40 \end{cases}$$

Upper plot: the discrete sequence $u(t)$

Bottom plot: the continuous signal u_{cont} made by the ZOH (red) compared with the discrete sequence $u(t)$ (blue)

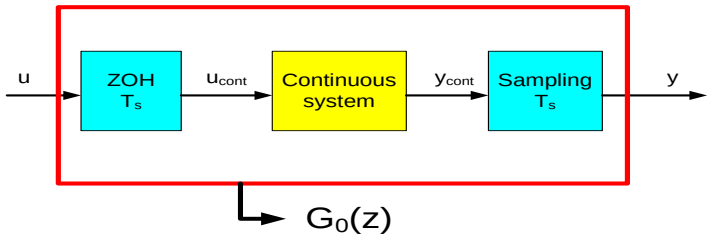


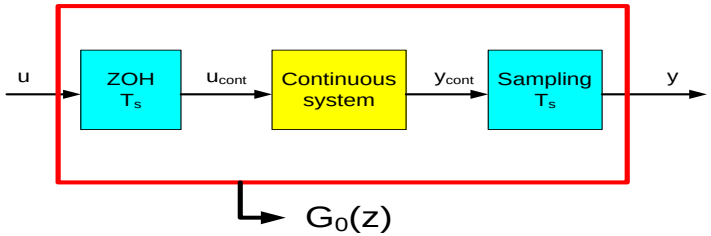
The continuous signal u_{cont} is then filtered by $G_0(s)$ delivering the continuous signal y_{cont} (upper plot, red). This continuous signal is then sampled with a sample period $T_s = 0.04s$. (upper plot, blue circle). This delivers the discrete sequence $y(t)$ of 41 samples ($t = 0...40$) (bottom plot)



Discrete-time transfer function

Does it exist a transfer function relation between $y(t)$ and $u(t)$?





$$G_0(z) \triangleq \frac{Y(z)}{U(z)} = \frac{\sum_{t=-\infty}^{+\infty} y(t)z^{-t}}{\sum_{t=-\infty}^{+\infty} u(t)z^{-t}}$$

$G_0(z)$ can be computed from $G_0(s)$ with the function $c2d.m$ of Matlab (ZOH methodology)

Example:

When $G_0(s) = \frac{a}{s+a}$ and $T_s = 0.04s.$, the discrete-time transfer function between $y(t)$ and $u(t)$ is

$$G_0(z) = \frac{(1-b)z^{-1}}{1-bz^{-1}} \quad \text{with } b = e^{-aT_s}$$

Thus:

$$G_0(s) = \frac{10}{s+10} \longleftrightarrow G_0(z) = \frac{0.33z^{-1}}{1-0.67z^{-1}}$$

Proof:

Suppose $u(t)$ is a discrete step, then $u_{cont}(t_c)$ is a continuous step. The step response of $G_0(s)$ is, for $t_c > 0$,

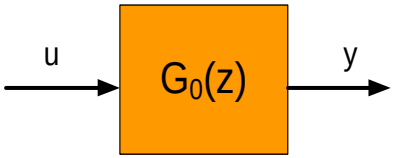
$$y_{cont}(t_c) = 1 - e^{-at_c}$$

The sampled signal $y(t)$ is given by $y_{cont}(t_c)$ at samples $t_c = tT_s$ i.e., for $t > 0$,

$$\begin{aligned} y(t) \triangleq y_{cont}(tT_s) &= 1 - e^{-atT_s} \\ &= 1 - b^t \end{aligned}$$

$$G_0(z) = \frac{Y(z)}{U(z)} = \frac{\frac{1}{1-z^{-1}} - \frac{1}{1-bz^{-1}}}{\frac{1}{1-z^{-1}}} = \frac{(1-b)z^{-1}}{1-bz^{-1}}$$

Properties of discrete-time transfer function



With some abuse, we will write

$$y(t) = G_0(z)u(t)$$

$$y(t) = G_0(z)u(t)$$

can be seen as a difference equation since:

$$z^{-1}u(t) \triangleq u(t-1)$$

Example:

$$y(t) = \frac{bz^{-1}}{1-az^{-1}}u(t) \Longleftrightarrow y(t) - ay(t-1) = bu(t-1)$$

this allows to compute the sequence $y(t)$ as a function of the sequence $u(t)$

Remark:

pure delays can be easily represented within $G_0(z)$

For continuous transfer function, a pure delay of $\alpha = \beta T_s$ seconds (β integer) is a non-rational part:

$$e^{-\alpha s}\frac{10}{s+10}$$

The corresponding rational discrete transfer function is:

$$z^{-\beta}\frac{0.33z^{-1}}{1-0.67z^{-1}}$$

Impulse response of $G_0(z)$

Assume $G_0(z)$ is causal

The impulse response $g_0(t)$ $t = 0..+\infty$ is the response $y(t) = G_0(z)u(t)$ when $u(t)$ is a discrete pulse $\delta(t)$ i.e. $u(t) = 1$ when $t = 0$ and $u(t) = 0$ elsewhere

This response allows to rewrite $G_0(z)$ as follows:

$$G_0(z) = \sum_{k=0}^{\infty} g_0(k)z^{-k}$$

Indeed:

$$y(t) = G_0(z)\delta(t) = \sum_{k=0}^{\infty} g_0(k)\delta(t-k) = g_0(t)$$

The impulse sequence $g_0(t)$ can be deduced

- by solving the difference equation for $u(t) = \delta(t)$
- by dividing the numerator of $G_0(z)$ by its denominator

Stability of $G_0(z)$

a transfer function is stable $\Longleftrightarrow$ the poles of $G_0(z)$ are all located within the unit circle

Example:

$$\frac{bz^{-1}}{1-az^{-1}} \quad stable \Longleftrightarrow |a| < 1$$

Indeed, $z = a$ is the unique pole of $1 - az^{-1}$

Frequency response of $G_0(z)$

the frequency response of $G_0(z)$ is given by the transfer function evaluated at $z = e^{j\omega}$ i.e. on the unit-circle:

$$G_0(z = e^{j\omega})$$

Only the frequency response between $[0\ \pi]$ is relevant.

Discrete frequency $\omega \in [0\ \pi] \Longrightarrow$ actual frequency $\omega_{actual} = \frac{\omega}{T_s}$ (ω_{actual} lies within the interval between 0 and the Nyquist pulsation)

General interpretation:

$$Y(\omega) = G_0(e^{j\omega})U(\omega)$$

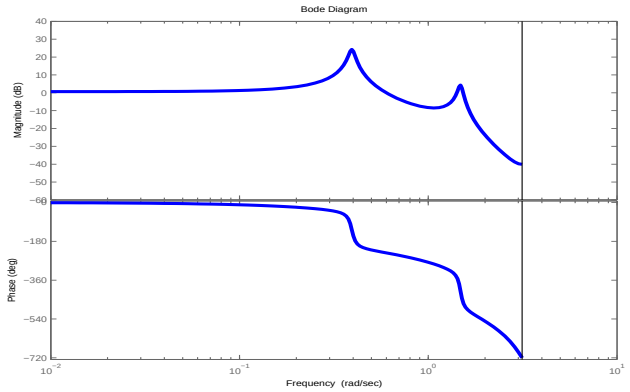
with $Y(\omega)$, $U(\omega)$ the Fourier transform of $y(t)$, $u(t)$ ($t = -\infty ... + \infty$)

One particular consequence:

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

$$y(t) = G_0(z)u(t) = |G_0(e^{j\omega_0})|\ sin\left(\omega_0 t + \angle G_0(e^{j\omega_0})\right)$$

Frequency response representation: bode plot



Remarks

1. Choice of T_s

The sampling period T_s is an important variable

It should be chosen so that $[0\ \frac{\pi}{T_s}]$ covers the band of significance of the continuous-time system

See end of the course for methodologies to choose T_s

2. Non-linearities

We adopt a linear framework to define the relation between u and y

We thus analyze the behaviour around one particular set-point

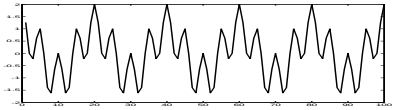
If the system is used at multiple set-points, a model must be identified for each of them (and coupled with a scheduling function)

3 Discrete-time signal analysis

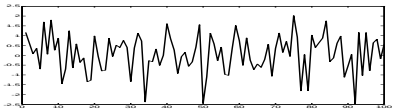
Signals encountered in system identification

Input $u(t)$:

multisine

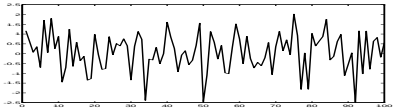


(filtered)
white noise



Disturbance $v(t)$:

stochastic
signal



Output $y(t)$:

$$y(t) = G_0(z)u(t) + v(t)$$

Observations

finite-power signals $\Longrightarrow$ analysis via their power spectrum $\Phi(\omega)$ (i.e. distribution of power content over the frequency ω)

signal $y(t)$ can be made up of a combination of stochastic and deterministic signal (e.g. when $u(t)$ is a multisine)

$\Longrightarrow$ make it complicate to define $\Phi(\omega)$

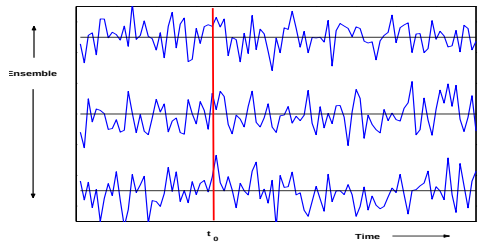
A new theory is necessary to deal with such signals called quasi-stationary signals (see later)

Recap: Stochastic vs. Deterministic signals

the values taken by a stationary stochastic signal at different t are different at each experiment/realization

BUT, each realization has the same power content over ω (i.e. the same $\Phi(\omega)$)

three realizations



Stationarity also implies that the mean of the signal and the auto-correlation function is time-invariant

the values taken by a deterministic signal at different t and thus $\Phi(\omega)$ are the same for all experiments/realizations

In identification, the deterministic signals are the multisines

Analysis of quasi-stationary signals

A quasi-stationary signal is a finite-power signal which can be

- a stochastic signal (stationary)
- a deterministic signal
- the summation of a stochastic and a deterministic signal

Analysis very close to the one of stationary stochastic signals (see WB2310 S&R3)

Mean $\bar{E}u(t)$ of a quasi-stationary signal $u(t)$

Mean of a deterministic signal $u(t)$: $\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} u(t)$

Mean of a stochastic signal $u(t)$: $Eu(t)$

$\Longrightarrow$ New operator $\bar{E}$

$$\bar{E}u(t) \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} Eu(t)$$

for purely stochastic or deterministic signal, the new operator is equivalent to the classical mean operator

Power spectrum $\Phi_u(\omega)$ of a quasi-stationary signal

The power spectrum of $u(t)$ is defined as the Fourier Transform of the auto-correlation function of $u(t)$:

$$\Phi_u(\omega) \triangleq \sum_{\tau=-\infty}^{+\infty} R_u(\tau)\ e^{-j\omega\tau}$$

with

$$R_u(\tau) \triangleq \bar{E}\left(u(t)\ u(t-\tau)\right)$$

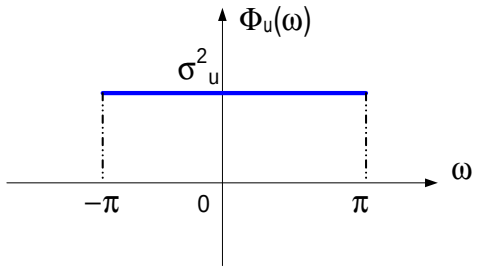
Total power $\mathcal{P}_u \triangleq \bar{E}u^2(t)$ of $u(t)$:

$$\mathcal{P}_u = R_u(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_u(\omega)d\omega$$

Example 1: $\Phi_u(\omega)$ and $\mathcal{P}_u$ when $u(t)$ is a white noise of variance σ_u^2?

$$\begin{aligned} R_u(\tau) &= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} E\left(u(t)u(t-\tau)\right) \\ &= E\left(u(t)u(t-\tau)\right) \quad \text{by stationarity} \\ &\triangleq \begin{cases} \sigma_u^2 & \text{when } \tau = 0 \\ 0 & \text{when } \tau \neq 0 \end{cases} \end{aligned}$$

$$\Phi_u(\omega) = \sum_{\tau=-\infty}^{+\infty} R_u(\tau)e^{-j\omega\tau} = \sigma_u^2\, e^{-j\omega 0} = \sigma_u^2 \;\forall\omega$$



and $\mathcal{P}_u = R_u(0) = \sigma_u^2$

Example 2: $\Phi_u(\omega)$ and $\mathcal{P}_u$ when $u(t) = Asin(\omega_0 t + \phi)$

$$\begin{aligned} R_u(\tau) &= \bar{E}\left(u(t)u(t-\tau)\right) \\ &= \bar{E}\left(A^2\, sin(\omega_0 t + \phi)\; sin(\omega_0 t - \omega_0\tau + \phi)\right) \\ &= \bar{E}\left(\frac{A^2}{2}cos(\omega_0\tau) - \frac{A_0^2}{2}cos(2\omega_0 t - \omega_0\tau + 2\phi)\right) \end{aligned}$$

$$\Longrightarrow R_u(\tau) =$$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\left(\frac{A^2}{2}cos(\omega_0\tau) - \frac{A_0^2}{2}cos(2\omega_0 t - \omega_0\tau + 2\phi)\right)$$

since $Es(t) = s(t)$ for a deterministic signal.

$$\Longrightarrow R_u(\tau) = \frac{A^2}{2}cos(\omega_0\tau)$$

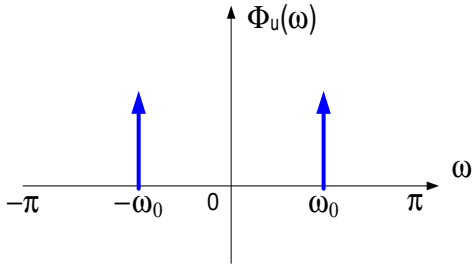
and thus, in the fundamental frequency range $[-\pi\ \pi]$,

$$\Phi_u(\omega) = \frac{A^2\pi}{2}\left(\delta(\omega-\omega_0) + \delta(\omega+\omega_0)\right)$$

and $\mathcal{P}_u = R_u(0) = \frac{A^2}{2}$.

$$\Phi_u(\omega) = \frac{A^2\pi}{2}\left(\delta(\omega-\omega_0)+\delta(\omega+\omega_0)\right)$$

The power spectrum of the sinus is independent of its phase shift ϕ and is = to 0 except in $\pm\omega_0$ where it is infinite.



Properties of the power spectrum

$$y(t) = G(z)u(t) \Longrightarrow \Phi_y(\omega) = |G(e^{j\omega})|^2\ \Phi_u(\omega)$$

$$y(t) = s_1(t) + s_2(t) \text{ with } s_1(t) \text{ independent of } s_2(t)$$

$$\Longrightarrow \Phi_y(\omega) = \Phi_{s_1}(\omega) + \Phi_{s_2}(\omega)$$

Cross- and auto-correlation function

The cross-correlation $R_{yu}(\tau)$ between y and u is a function which allows to verify whether two q-s signals $y(t)$ and $u(t)$ are correlated with each other

$$R_{yu}(\tau) \triangleq \bar{E}\left(y(t)u(t-\tau)\right)$$

Properties:

- the value of $y(t)$ at time t is not (cor)related in any way to the value of $u(t-\tau) \Longrightarrow R_{yu}(\tau) = 0$
- the signals $y(t)$ and $u(t)$ are independent $\Longrightarrow$ $R_{yu}(\tau) = 0\ \forall\tau$

NB. $R_u(\tau) = R_{uu}(\tau)$

Approximations of $R_u(\tau)$ and $\Phi_u(\omega)$ using finite data

To exactly compute $R_u(\tau)$ and $\Phi_u(\omega)$, we need both an infinite number of measurements of $u(t)$ and an infinite number of realizations of $u(t)$

In practice, we have generally $N < \infty$ measurements of $u(t)$: $\{u(t) \mid t = 0...N-1\}$

A. Approximation of $R_u(\tau)$ and properties of this approximation

$$\hat{R}_u^N(\tau) = \begin{cases} \frac{1}{N}\sum_{t=0}^{N-1} u(t)u(t-\tau) & for\ |\tau| < N-1 \\ 0 & for\ |\tau| > N-1 \end{cases}$$

This approximation is a consistent estimate of $R_u(\tau)$ i.e.

$$\lim_{N\to\infty} \hat{R}_u^N(\tau) = R_u(\tau)$$

For fixed N, though, the accuracy of $\hat{R}_u^N(\tau)$ decreases for increasing values of τ since $\hat{R}_u^N(\tau)$ is computed with lesser and lesser products $u(t)u(t-\tau)$

B. Approximation of $\Phi_u(\omega)$ (Periodogram) and properties of this approximation

$\Phi_u(\omega)$ can be approximated in two equivalent ways:

$$\begin{aligned}\hat{\Phi}_u^N(\omega) &= \sum_{\tau=-\infty}^{+\infty} \hat{R}_u^N(\tau)\, e^{-j\omega\tau} \\ &= U_N(\omega)\, U_N^*(\omega)\end{aligned}$$

with $U_N(\omega)$ the (scaled) Fourier Transform of $\{u(t) \mid t = 0...N-1\}$ i.e.

$$U_N(\omega) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} u(t)\, e^{-j\omega t}$$

Note: the approximation via $U_N(\omega)$ is the most logical for deterministic signals

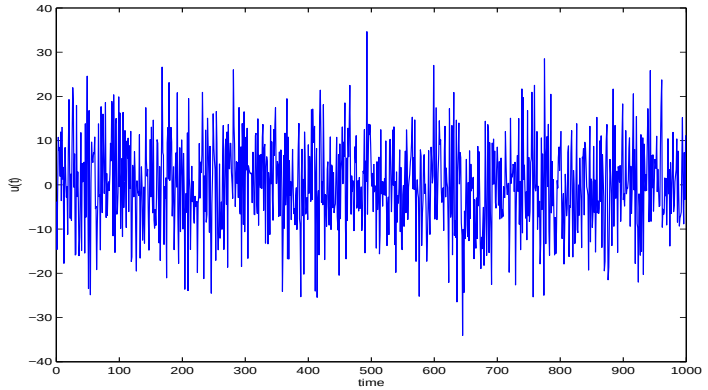
When $u(t)$ is deterministic, $\hat{\Phi}_u^N(\omega)$ is a consistent estimate of $\Phi_u(\omega)$

$$\lim_{N\to\infty} \hat{\Phi}_u^N(\omega) = \Phi_u(\omega)$$

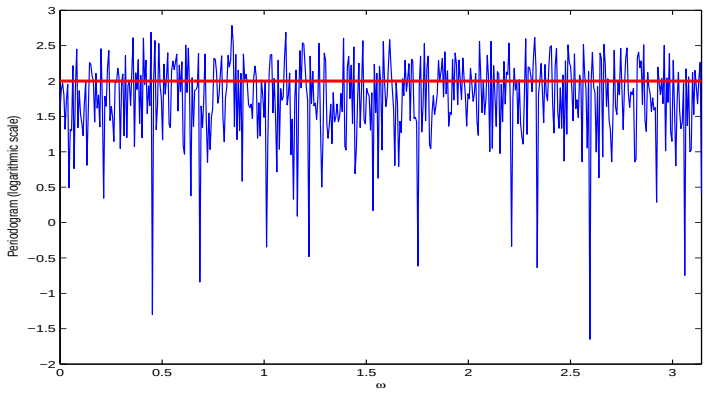
For all other cases, we have only that $\hat{\Phi}_u^N(\omega)$ is an asymptotically unbiased estimate of $\Phi_u(\omega)$ (variance is nonzero)

$$\lim_{N\to\infty} E\hat{\Phi}_u^N(\omega) = \Phi_u(\omega)$$

Example 1: we have collected $N = 1000$ time-samples of a white noise of variance $\sigma_u^2 = 100$

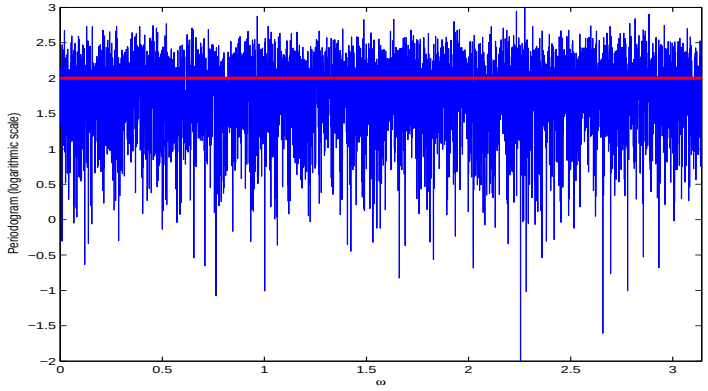


Obtained Periodogram $\hat{\Phi}_u^N(\omega)$ (blue) w.r.t. $\Phi_u(\omega)$ (red)

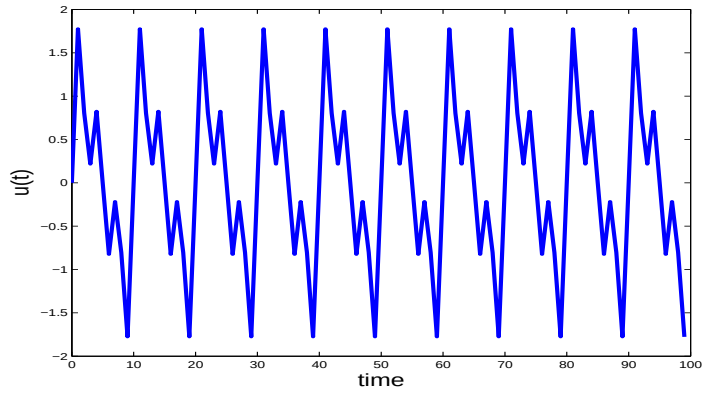


$\hat{\Phi}_u^N(\omega)$ **is an erratic function fluctuating around** $\Phi_u(\omega)$

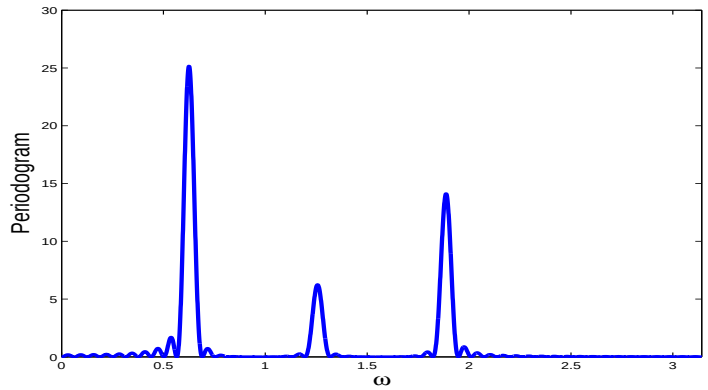
As expected, it does not change when N is increased to $N = 10000$:



Example 2: we have collected $N = 100$ time-samples of $u(t) = sin(0.63t) + \frac{1}{2}sin(1.26t) + \frac{3}{4}sin(1.89t)$ (fundamental period = 10 time-samples):

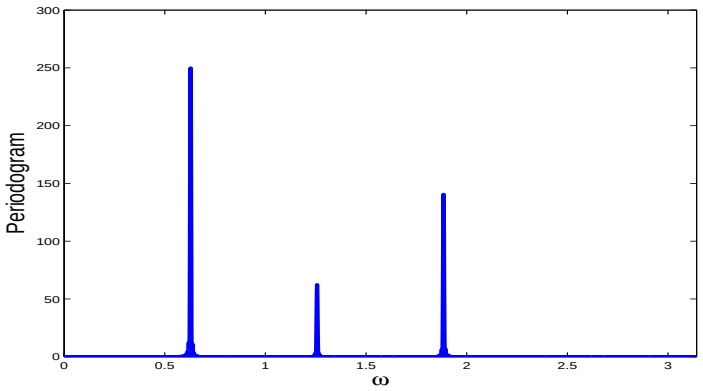


Obtained Periodogram $\hat{\Phi}_u^N(\omega)$ for $\omega \in [0\ \pi]$



It can be proven that the value at $\omega_1 = 0.63$, $\omega_2 = 1.26$ and $\omega_3 = 1.89$ are given by $\frac{NA_i^2}{4}$ where A_i $(i = 1, 2, 3)$ is the amplitude of the sinusoid of frequency ω_i.

For $N \to \infty$, $\hat{\Phi}_u^N(\omega)$ tends thus to $\Phi_u(\omega)$. Here is the periodogram for the same signal when $N = 1000$

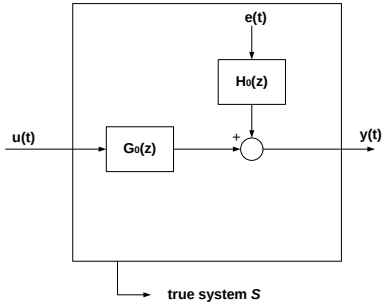


Part III: Prediction Error Identification

1. Introduction about Prediction Error Identification

1.1. Assumptions on the True System: $\mathcal{S} = \{ G_0 \quad H_0 \}$

$$y(t) = G_0(z)u(t) + \overbrace{H_0(z)e(t)}^{v(t)}$$



$G_0(z)$ and $H_0(z)$ are two unknown linear transfer functions in the $\mathcal{Z}$-transform (e.g. $G_0(z) = \frac{3z^{-1}}{1+0.5z^{-1}}$ and $H_0(z) = \frac{1}{1+0.5z^{-1}}$)

the input signal $u(t)$ is chosen by the operator and applied to $\mathcal{S}$ and the output signal $y(t)$ is measured

$y(t)$ is assumed to be made up of two distinct contributions:

- $G_0u(t)$: dependent of the choice of $u(t)$
- the disturbance $v(t) = H_0(z)e(t)$: independent of the input signal $u(t)$

the disturbance $v(t)$ represents the measurement noise; the effects of stochastic disturbance, the effects of non-measurable input signals; $\cdots$

the disturbance $v(t)$ is modeled by $H_0(z)e(t)$:

- $H_0(z)$ is stable, inversely stable and monic (i.e. $H_0(z) = 1 + \sum_{k=1}^{\infty} h_0(k)z^{-k}$)
- $e(t)$ is a white noise signal i.e. a sequence of independent, identically distributed random variables (no assumption is made on the probability density function)

Properties of $e(t)$ and $v(t)$ as a consequence of the assumptions

Since $\{e(t)\}$ is a white noise,

$$\begin{aligned} Ee(t) &= 0 \\ R_e(\tau) \triangleq Ee(t)e(t-\tau) &= \sigma_e^2 \cdot \delta(\tau) \end{aligned}$$

$\{v(t)\}$ is therefore the realization of a stochastic process with properties:

$$\begin{aligned} Ev(t) &= 0 \\ \Phi_v(\omega) &= |H_0(e^{i\omega})|^2 \cdot \sigma_e^2 \end{aligned}$$

1.2. Objective of PE Identification

General Objective

Find the best parametric models $G(z,\theta)$ and $H(z,\theta)$ for the unknown transfer functions G_0 and H_0 using a set of measured data $u(t)$ and $y(t)$ generated by the true system $\mathcal{S}$.

Example of parametric models:

$$G(z,\theta) = \frac{\theta_1 z^{-1}}{1+\theta_2 z^{-1}} \qquad H(z,\theta) = \frac{1}{1+\theta_2 z^{-1}}$$

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \qquad \mathcal{M} = \{\ G(z,\theta),\ H(z,\theta) \quad \forall \theta \in \mathbf{R}^2\ \}$$

Note: $H(z,\theta)$ is always chosen as a monic transfer function (like H_0)

In the beginning, we will make the following assumption:

$$\exists \theta_0 \ \ such\ that \ \ G(z,\theta_0) = G_0(z) \ \ and \ \ H(z,\theta_0) = H_0(z)$$

$$i.e. \ \ \mathcal{S} \in \mathcal{M}$$

The objective can therefore be restated as follows:

Find (an estimate of) the unknown parameter vector θ_0 using a set of N input and output data:

$$Z^N = \{\ u(t),\ y(t) \ \mid \ t = 1...N\ \}$$

generated by the true system i.e. $y(t) = G_0 u(t) + H_0 e(t)$

Summary: the full-order identification problem

Consider the following true system:

$$y(t) = G_0(z)u(t) + \overbrace{H_0(z)e(t)}^{v(t)} = G(z,\theta_0)u(t) + H(z,\theta_0)e(t)$$

from which N input and output data have been measured:

$$Z^N = \{\ u(t),\ y(t) \ \mid \ t = 1...N\ \}.$$

Given the parametrization $G(z,\theta)$ and $H(z,\theta)$, find (an estimate of) the unknown parameter θ_0.

Simple idea to reach this objective :

Let us simulate the parametric models with the input $u(t)$ in Z^N:

$$y(t,\theta) = G(z,\theta)u(t) + H(z,\theta)e(t)$$

and let us find the vector θ for which:

$$y(t) - y(t,\theta) = 0 \quad \forall t = 1...N$$

In other words, $\theta = \theta_0$ minimizes the power of $y(t) - y(t,\theta)$

Problem: $y(t,\theta)$ can not be computed since the white noise sequence $e(t)$ is unknown

Consequences:

- we need to find an accurate way to *predict* $y(t,\theta)$
- the predictor $\hat{y}(t,\theta)$ should be chosen in such a way that θ_0 can still be deduced e.g. by minimizing the power of $y(t) - \hat{y}(t,\theta)$

2. Predictor $\hat{y}(t,\theta)$ in identification and prediction error $\epsilon(t,\theta)$

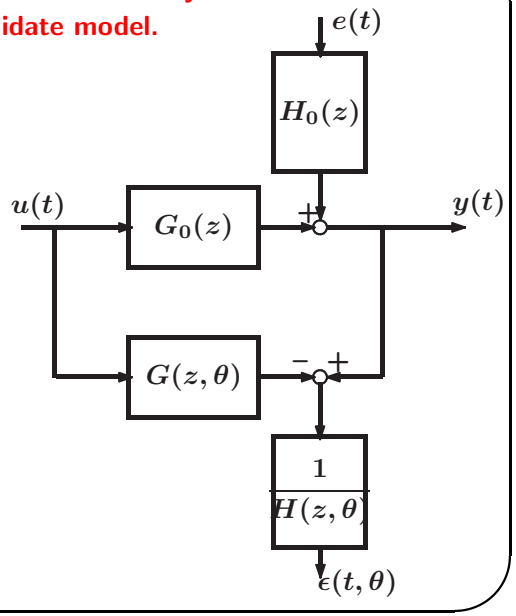
Given Z^N and a model $G(z,\theta)$, $H(z,\theta)$ in $\mathcal{M}$, we define the predictor $\hat{y}(t,\theta)$ of the output of this model as follows:

$$\hat{y}(t,\theta) \triangleq H(z,\theta)^{-1}G(z,\theta)u(t) + (1 - H(z,\theta)^{-1})y(t) \; \forall t = 1...N$$

and we define the prediction error $\epsilon(t,\theta)$ as follows:

$$\begin{aligned} \epsilon(t,\theta) &\triangleq y(t) - \hat{y}(t,\theta) \quad \forall t = 1...N \\ &= H(z,\theta)^{-1}\left(y(t) - G(z,\theta)u(t)\right) \quad \forall t = 1...N \end{aligned}$$

$\epsilon(t,\theta)$ compares the output of the true system and the predicted output of a candidate model.



Properties of the prediction error $\epsilon(t,\theta)$

Property 1. Given θ and Z^N, $\epsilon(t,\theta)$ computable $\forall t = 1...N$

Example:

$$G(z,\theta) = \frac{\theta_1 z^{-1}}{1+\theta_2 z^{-1}} \quad H(z,\theta) = \frac{1}{1+\theta_2 z^{-1}} \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$$

$$\begin{aligned} \epsilon(t,\theta) &= (1+\theta_2 z^{-1})\left(y(t) - \frac{\theta_1 z^{-1}}{1+\theta_2 z^{-1}} u(t)\right) \\ &= y(t) + \theta_2 y(t-1) + \theta_1 u(t-1) \end{aligned}$$

Notes:

it is typically assumed that $u(t<0) = y(t<0) = 0$

$H^{-1}(z,\theta)$ is always causal since $H(z,\theta)$ is monic !

Property 2. $\epsilon(t,\theta_0) = e(t)$ (smth really unpredictable at time t)

$$\begin{aligned} \epsilon(t,\theta) &= H(z,\theta)^{-1}\left(\overbrace{G_0(z)u(t) + H_0(z)e(t)}^{y(t)} - G(z,\theta)u(t)\right) \\ &= \frac{G_0(z) - G(z,\theta)}{H(z,\theta)} u(t) + \frac{H_0}{H(z,\theta)} e(t) \end{aligned}$$

$$\Longrightarrow \epsilon(t,\theta_0) = e(t)$$

Property 3. $\epsilon(t,\theta) \neq$ white noise for all $\theta \neq \theta_0$ (provided an appropriate signal $u(t)$)

Property 4. θ_0 minimizes the power $\bar{E}\epsilon^2(t,\theta)$ of $\epsilon(\theta)$ i.e.

$$\theta_0 = arg\min_\theta \bar{E}\epsilon^2(t,\theta)$$

$$with\ \bar{E}\epsilon^2(t,\theta) \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} E\epsilon^2(t,\theta)$$

Since $\epsilon(t,\theta_0) = e(t)$, we have thus:

$$\bar{E}\epsilon^2(t,\theta_0) = \sigma_e^2$$
$$\bar{E}\epsilon^2(t,\theta) > \sigma_e^2 \quad \forall \theta \neq \theta_0$$

(the latter provided $u(t)$ has been chosen appropriately)

Sketch of the proof of Property 4:

$$\epsilon(t,\theta) = e(t) + \overbrace{\frac{G_0(z) - G(z,\theta)}{H(z,\theta)} u(t)}^{s_1(t,\theta)} + \overbrace{\frac{H_0(z) - H(z,\theta)}{H(z,\theta)} e(t)}^{s_2(t,\theta)}$$

with $s_2(t,\theta)$ function of $e(t-1)$, $e(t-2)$, ... (not of e(t)).

$u(t)$ and $e(t)$ uncorrelated and $e(t)$ white noise $\Longrightarrow$

$$\bar{E}\epsilon^2(t,\theta) = \sigma_e^2 + \bar{E}s_1^2(t,\theta) + \bar{E}s_2^2(t,\theta)$$

$\theta = \theta_0$ minimizes both $\bar{E}s_1^2(t,\theta)$ and $\bar{E}s_2^2(t,\theta)$ by making them equal to 0.

$\Longrightarrow \theta = \theta_0$ minimizes $\bar{E}\epsilon^2(t,\theta)$

Important remark. The two following statements are equivalent:

- The true parameter vector θ_0 reduces the prediction error $\epsilon(t,\theta)$ to the realization of the noise $e(t)$.
- The true parameter vector θ_0 minimizes the power of the prediction error $\epsilon(t,\theta)$.

Example

We have collected $N = 2000$ data $u(t)$ and $y(t)$ on the following true system

$$y(t) = \frac{z^{-3}\left(0.103 + 0.181z^{-1}\right)}{1 - 1.991z^{-1} + 2.203z^{-2} - 1.841z^{-3} + 0.894z^{-4}}u(t) + e(t)$$

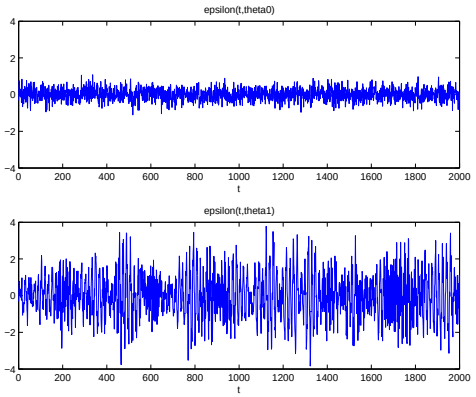
and we have chosen the following model structure $\mathcal{M}$

$$\mathcal{M} = \left\{ G(z,\theta) = \frac{z^{-3}\left(b_0 + b_1 z^{-1}\right)}{1 + f_1 z^{-1} + f_2 z^{-2} + f_3 z^{-3} + f_4 z^{-4}} \; ; \; H(z,\theta) = 1 \right\}$$

$$\theta = \left(\; b_0 \; , \; b_1 \; , \; f_1 \; , \; f_2 \; , \; f_3 \; , \; f_4 \; \right)^T$$

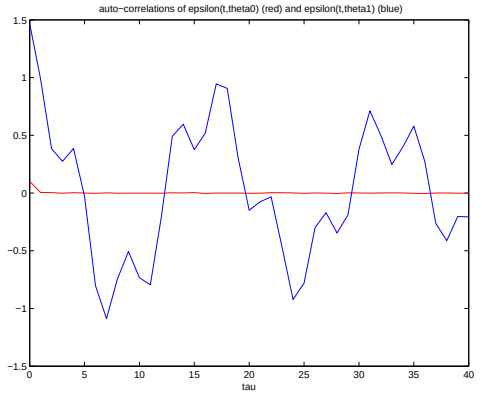
$$\Longrightarrow \theta_0 = \left(\; 0.103 \; , \; 0.181 \; , \; -1.991 \; , \; 2.203 \; , \; -1.841 \; , \; 0.894 \; \right)^T$$

We have computed $\epsilon(t,\theta)$ $(t = 1...N)$ for $\theta = \theta_0$ and for another θ i.e. $\theta_1 \neq \theta_0$:



$$\theta_1 = \left(\; 0.12 \; , \; 0.25 \; , \; -2 \; , \; 2.3 \; , \; -1.9 \; , \; 0.8 \; \right)^T$$

As can be seen with $\hat{R}_{\epsilon}^{N}(\tau)$, $\epsilon(t,\theta_0)$ has well the properties of a white noise as opposed to $\epsilon(t,\theta_1)$



Estimated power of $\epsilon(t, \theta_0)$: 0.1015 $(\sigma_e^2 = 0.1)$

Estimated power of $\epsilon(t, \theta_1)$: 1.4678

Note: the estimated power is $\hat{R}_\epsilon^N(0)$

Summary:

- $\epsilon(t, \theta)$ is a computable quantity comparing the output $y(t)$ of the true system and the predicted output of a model
- $\theta = \theta_0$ minimizes the power of $\epsilon(t, \theta)$

3. Mathematical criterion for prediction error identification

3.1. An ideal criterion

Denote by θ^*, the solution of the minimization of the power of the prediction error:

$$\theta^* = arg\min_\theta \bar{V}(\theta)$$

$$with\ \bar{V}(\theta) = \bar{E}\epsilon^2(t, \theta) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} E\epsilon^2(t, \theta)$$

Properties of $\bar{V}(\theta)$ and θ^* (when $\mathcal{S} \in \mathcal{M}$ and $u(t)$ appropriate)

$\bar{V}(\theta)$ has an unique minimum θ^*

$$\theta^* = \theta_0$$

Remark:

There is no difference between θ^* and θ_0 at this stage of the course since we suppose $\mathcal{S} \in \mathcal{M}$ and $u(t)$ appropriate.

We nevertheless introduce the new notation θ^* since

- when $\mathcal{S} \notin \mathcal{M}$, the notion of true parameter vector θ_0 does not exist, while the minimum θ^* of the cost function $\bar{V}(\theta)$ exists
- if $u(t)$ is not chosen appropriately, then $\bar{V}(\theta)$ has several minima and θ^* represents the set of these minima, while θ_0 is one single parameter vector

The true parameter vector θ_0 is thus the solution of:

$$arg\min_\theta \bar{V}(\theta)$$

$$with\ \bar{V}(\theta) = \bar{E}\epsilon^2(t,\theta) = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} E\epsilon^2(t,\theta)$$

Question ? Is it possible to consider this criterion ? NO !!!

Indeed, the power of the prediction error can not be exactly computed with only one experiment and only N measured data.

3.2. Tractable identification criterion

Power of prediction error is estimated using the N available data Z^N:

$$\begin{aligned} V_N(\theta, Z^N) &= \frac{1}{N}\sum_{t=1}^{N}\epsilon^2(t,\theta) \\ &= \frac{1}{N}\sum_{t=1}^{N}\left((H(\theta)^{-1}(y(t) - G(\theta)u(t))\right)^2 \end{aligned}$$

Parameter estimation through minimizing V_N:

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

Consequences and properties of the identified parameter vector $\hat{\theta}_N$:

- different experiments and data $\Longrightarrow$ different $\hat{\theta}_N$.
- $\hat{\theta}_N$ is only an estimate of $\theta^*(=\theta_0)$.
- $\hat{\theta}_N$ is a random variable which is asymptotically ($N\to\infty$) Gaussian with mean θ^*:

$$\hat{\theta}_N \sim As\mathcal{N}(\theta^*, P_\theta)$$

- $\hat{\theta}_N \to \theta^*$ with probability 1 when $N\to\infty$ (i.e. $P_\theta \to 0$ when $N\to\infty$)

Example:

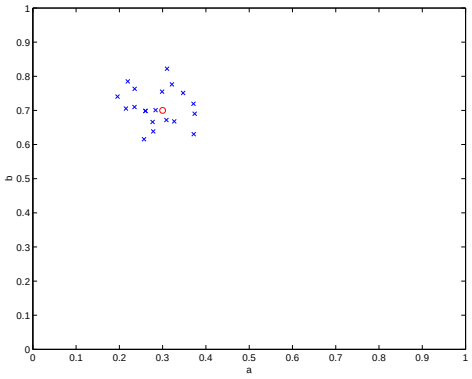
$$\mathcal{S}: \quad y(t) = \frac{0.7z^{-1}}{1+0.3z^{-1}}u(t) + \frac{1}{1+0.3z^{-1}}e(t)$$

$$\mathcal{M}: \quad G(z,\theta) = \frac{bz^{-1}}{1+az^{-1}} \quad H(z,\theta) = \frac{1}{1+az^{-1}} \quad \theta = \begin{pmatrix} a \\ b \end{pmatrix}$$

we have applied 20 times the same sequence $u(t)$ of length $N = 200$ and we have measured the corresponding $y(t)$.

For these 20 experiments, we have computed the estimate $\hat{\theta}_N$ of $\theta_0 = \begin{pmatrix} 0.3 \ , \ 0.7 \end{pmatrix}^T$

The twenty estimates $\hat{\theta}_N$ are represented with a blue cross and θ_0 by a red circle.



How can we solve the optimization problem delivering $\hat{\theta}_N$?

$$\begin{aligned}\hat{\theta}_N &= \arg\min_{\theta} \frac{1}{N}\sum_{t=1}^{N} \epsilon^2(t,\theta) \\ &= \arg\min_{\theta} \frac{1}{N}\sum_{t=1}^{N} \left((H(\theta)^{-1}(y(t) - G(\theta)u(t))\right)^2\end{aligned}$$

In order to answer this question, the parametrization of $G(z,\theta)$ and $H(z,\theta)$ must be defined more precisely.

4 Black box model structures

Model structure: $\mathcal{M} = \{(G(z,\theta), H(z,\theta)), \theta \in \mathbb{R}^{n_\theta}\}$

General parametrization used in the Matlab Toolbox:

$$G(z,\theta) = \frac{z^{-n_k}B(z,\theta)}{F(z,\theta)A(z,\theta)} \quad H(z,\theta) = \frac{C(z,\theta)}{D(z,\theta)A(z,\theta)}$$

$$\theta^T = \begin{pmatrix} a_1 & .. & a_{n_a} & b_0 & .. & f_{n_f} \end{pmatrix}$$

$$B(z,\theta) = b_0 + b_1 z^{-1} + \cdots + b_{n_b-1} z^{-n_b+1}$$
$$A(z,\theta) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a}$$
$$C(z,\theta) = 1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}$$
$$D(z,\theta) = 1 + d_1 z^{-1} + \cdots + d_{n_d} z^{-n_d}$$
$$F(z,\theta) = 1 + f_1 z^{-1} + \cdots + f_{n_f} z^{-n_f}$$

Model structures used in practice

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

Example: ARX Model structure

$$G(z,\theta) = \frac{z^{-n_k}B(z,\theta)}{A(z,\theta)}; \qquad H(z,\theta) = \frac{1}{A(z,\theta)}$$

with

$$\begin{aligned} B(z,\theta) &= b_0 + b_1 z^{-1} + \cdots + b_{n_b-1} z^{-n_b+1} \\ A(z,\theta) &= 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a} \\ \theta &= \left(\begin{array}{cccccc} a_1 & a_2 & \cdots a_{n_a} & b_0 & b_1 & \cdots & b_{n_b-1} \end{array} \right)^T . \end{aligned}$$

n_a, n_b are the number of parameters in the A and B polynomial.
n_k number of time delays

Distinction between model structures

- ARX and FIR have a predictor linear in θ

$$\begin{aligned} \hat{y}(t,\theta) &= z^{-n_k}B(\theta)u(t) + (1 - A(\theta))y(t) \\ &= \phi^T(t)\theta \end{aligned}$$

  is a linear function in $\theta \Rightarrow$ Important computational advantages.

- BJ, FIR and OE have an independent parametrization of $G(z,\theta)$ en $H(z,\theta)$

  There are no common parameters in G and H.
  $\Rightarrow$ Advantages for independent identification of G and H.

5 Computation of the identified parameter vector $\hat{\theta}_N$

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

5.1 Case of a predictor linear in θ (ARX and FIR)

$$G(\theta) = \frac{z^{-n_k}B(\theta)}{A(\theta)}; \qquad H(\theta) = \frac{1}{A(\theta)}$$

Predictor $\hat{y}(t,\theta)$:

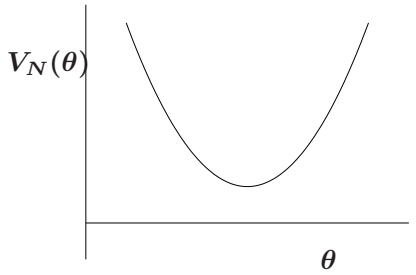
$$\begin{aligned} \hat{y}(t,\theta) &= H(\theta)^{-1}G(\theta)u(t) + [1 - H(\theta)^{-1}]y(t) \\ &= z^{-n_k}B(\theta)u(t) + [1 - A(\theta)]y(t) \\ &= \phi(t)^T\theta \quad \textbf{LINEAR in } \theta \textbf{ !!!} \end{aligned}$$

with

$$\phi(t) = (-y(t-1), ..., -y(t-n_a), \\ u(t-n_k), ..., u(t-n_k-n_b+1))^T$$

$$\theta = \begin{pmatrix} a_1 & a_2 & \cdots a_{n_a} & b_0 & \cdots b_{n_b-1} \end{pmatrix}^T$$

$V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} \left(y(t) - \phi^T(t)\theta\right)^2$ **is quadratic in** θ.



$\hat{\theta}_N = \arg\min_\theta \overbrace{\frac{1}{N}\sum_{t=1}^{N} \left(y(t) - \phi(t)^T\theta\right)^2}^{V_N(\theta,Z^N)}$ **can be determined analytically using:**

$$\left.\frac{\partial V_N(\theta, Z^N)}{\partial \theta}\right|_{\theta=\hat{\theta}_N} = 0$$

Indeed:

$$\frac{\partial V_N(\theta, Z^N)}{\partial \theta} = -2\frac{1}{N}\sum_{t=1}^{N}[\phi(t)y(t) - \phi(t)\phi^T(t)\theta]$$

Putting derivative to 0 in $\theta = \hat{\theta}_N$ delivers:

$$\left[\frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi^T(t)\right]\hat{\theta}_N = \frac{1}{N}\sum_{t=1}^{N}\phi(t)y(t)$$

As a consequence:

$$\hat{\theta}_N = \left[\underbrace{\frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi^T(t)}_{R(N)}\right]^{-1} \cdot \underbrace{\frac{1}{N}\sum_{t=1}^{N}\phi(t)y(t)}_{f(N)}$$

- **Analytical solution through simple matrix operations.**

5.2 Case of a predictor nonlinear in θ (OE,BJ,ARMAX)

Example of the OE model structure:

$$G(\theta) = \frac{z^{-n_k}B(\theta)}{F(\theta)}; \qquad H(\theta) = 1$$

Predictor $\hat{y}(t,\theta)$:

$$\begin{aligned}\hat{y}(t,\theta) &= H(\theta)^{-1}G(\theta)u(t) + [1 - H(\theta)^{-1}]y(t) \\ &= z^{-n_k}\frac{B(\theta)}{F(\theta)}u(t) \\ &= \phi(t,\theta)^T\theta \quad \text{NONLINEAR in } \theta \text{ !!!}\end{aligned}$$

with

$$\phi(t,\theta) = (u(t-n_k), ..., u(t-n_k-n_b+1), \\ -\hat{y}(t-1,\theta), ..., -\hat{y}(t-n_f,\theta))^T$$

$$\theta = \begin{pmatrix} b_0 & \cdots b_{n_b-1} & f_1 & f_2 \cdots f_{n_f} \end{pmatrix}^T$$

$$\hat{\theta}_N = \arg\min_\theta \overbrace{\frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \phi(t,\theta)^T\theta\right)^2}^{V_N(\theta,Z^N)}$$ **can not be determined analytically using:**

$$\left.\frac{\partial V_N(\theta, Z^N)}{\partial\theta}\right|_{\theta=\hat{\theta}_N} = 0$$

since this derivative is a very complicate expression which is nonlinear in θ and since this derivative is (generally) equal to 0 for several θ (local minima).

The solution $\hat{\theta}_N$ will therefore be computed iteratively. Risk of finding a local minimum !

6 Conditions on experimental data

The ideal identification criterion

$$arg\min_\theta \bar{E}\epsilon^2(t,\theta)$$

has a unique solution θ^* (i.e. $\theta^* = \theta_0$ when $\mathcal{S} \in \mathcal{M}$) if the input signal $u(t)$ that is chosen to generate the experimental data is sufficiently rich.

Counterexample

$$\mathcal{S}: y(t) = \frac{b_0 z^{-1}}{1 + f_0 z^{-1}}u(t) + \frac{1}{1 + d_0 z^{-1}}e(t)$$

Consider $u(t) = 0\ \forall t$ as input signal and a full-order model structure $\mathcal{M}$ for $\mathcal{S}$:

$$\mathcal{M} = \left\{ G(z,\theta) = \frac{bz^{-1}}{1+fz^{-1}};\ H(z,\theta) = \frac{1}{1+dz^{-1}} \quad \theta = \begin{pmatrix} b \\ d \\ f \end{pmatrix} \right\}$$

Consequently:

$$\epsilon(t,\theta) = \overbrace{\frac{G_0(z) - G(z,\theta)}{H(z,\theta)}u(t)}^{=0} + \frac{H_0}{H(z,\theta)}e(t)$$

$$\Longrightarrow \epsilon(t,\theta) = \frac{1+dz^{-1}}{1+d_0 z^{-1}} e(t)$$

We know that $\bar{E}\epsilon^2(t,\theta)$ is minimum for θ making $\epsilon(t,\theta) = e(t)$

$$\Longrightarrow$$

The power $\bar{E}\epsilon^2(t,\theta)$ is minimized for each θ making $H(z,\theta) = H_0$ i.e.

$$\theta^* = \left\{ \begin{pmatrix} b \\ d_0 \\ f \end{pmatrix} \quad \forall b \in \mathbf{R} \text{ and } \forall f \in \mathbf{R} \right\}$$

Note: θ_0 lies in the set of θ^*.

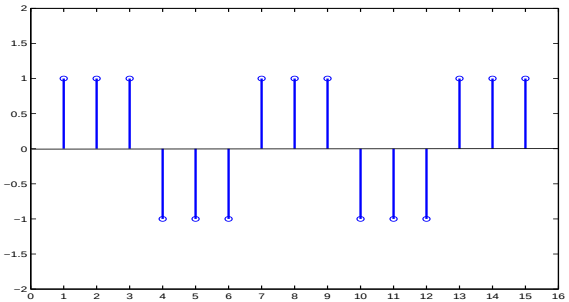
Notion of signal richness: persistently exciting input signals

A quasi-stationary signal u is persistently exciting of order n if the (Toeplitz) matrix $\bar{R}_n$ is non-singular

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

Examples:

- A white noise process ($R_u(\tau) = \sigma_u^2 \delta(\tau)$) is persistently exciting of infinite order. Indeed, $\bar{R}_n = \sigma_u^2 I_n$.
- a block signal



$$R_u(0) = 1 \quad R_u(1) = \tfrac{1}{3} \quad R_u(2) = -\tfrac{1}{3} \quad R_u(3) = -1$$
$$R_u(4) = -\tfrac{1}{3} \quad R_u(5) = \tfrac{1}{3} \quad R_u(6) = 1 \quad \text{etcetera}$$

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

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

Another method to determine the order of u

If the spectrum Φ_u is unequal to 0 in n points in the interval $(-\pi, \pi]$, then u is persistently exciting of order n

Example

The signal

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

is persistently exciting of order 2. (Φ_u has a contribution in $\pm\omega_0$).

Important result. Let us denote the number of parameters in the function $G(z, \theta)$ by n_g. The ideal identification criterion i.e.

$$\theta^* = arg \min_{\theta} \bar{V}(\theta)$$

has a unique solution (i.e. $\theta^* = \theta_0$) if the signal $u(t)$ generating the data is sufficiently exciting of order $\geq n_g$.

Sketch of the proof (case of a FIR model structure):

$$\epsilon(t, \theta) = y(t) - \sum_{k=1}^{n_b} b_k u(t-k) \quad (n_k = 1)$$

θ^* is characterized by:

$$\begin{bmatrix} R_u(0) & \cdots & R_u(n_b-1) \\ R_u(1) & \cdots & R_u(n_b-2) \\ \vdots & \ddots & \vdots \\ R_u(n_b-1) & \cdots & R_u(0) \end{bmatrix} \overbrace{\begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_{n_b}^* \end{bmatrix}}^{\theta^*} = \begin{bmatrix} R_{yu}(1) \\ R_{yu}(2) \\ \vdots \\ R_{yu}(n_b) \end{bmatrix}$$

Consequence:

θ^* can uniquely be identified *if and only if* u is persistently exciting of order $\geq n_b$.

What can we say about the identification of $\hat{\theta}_N$?

$\hat{\theta}_N$ will be the (consistent) estimate of $\theta^* = \theta_0$ (the unique solution of the ideal criterion) if the input signal is sufficiently exciting of order $\geq n_g$.

Remark. In the sequel, we will always assume that the signal $u(t)$ has been chosen such that it is persistently exciting of sufficient order.

Example

Let us consider the following true system $\mathcal{S}$:

$$y(t) = \frac{z^{-3}\left(0.103 + 0.181z^{-1}\right)}{1 - 1.991z^{-1} + 2.203z^{-2} - 1.841z^{-3} + 0.894z^{-4}}u(t) + e(t)$$

we have chosen the full-order model structure $\mathcal{M}$

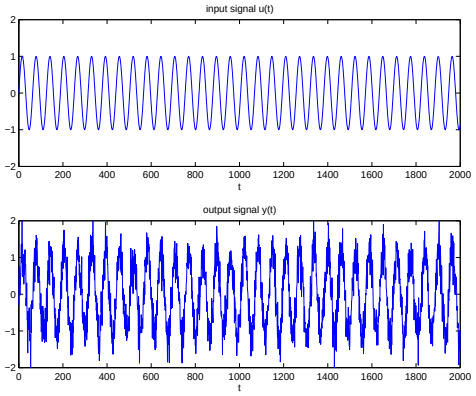
$$\mathcal{M} = \left\{ G(z,\theta) = \frac{z^{-3}\left(b_0 + b_1 z^{-1}\right)}{1 + f_1 z^{-1} + f_2 z^{-2} + f_3 z^{-3} + f_4 z^{-4}} \; ; \; H(z,\theta) = 1 \right\}$$

$$\theta = \left(\; b_0 \; , \; b_1 \; , \; f_1 \; , \; f_2 \; , \; f_3 \; , \; f_4 \; \right)^T \Longrightarrow n_G = 6$$

we now perform two identification experiments on $\mathcal{S}$

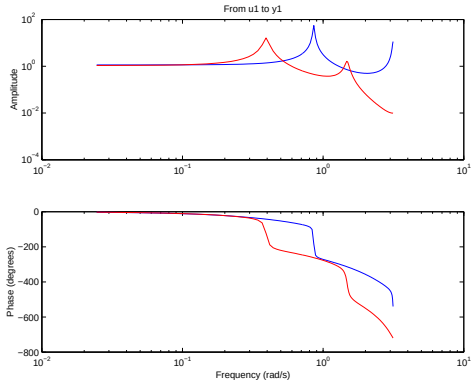
First experiment on $\mathcal{S}$

we have applied $u(t) = sin(0.1t)$ (u p.e. of order 2) to $\mathcal{S}$ and collected $N = 2000$ IO data:



Using the 2000 recorded data, we have identified $\hat{\theta}_N$

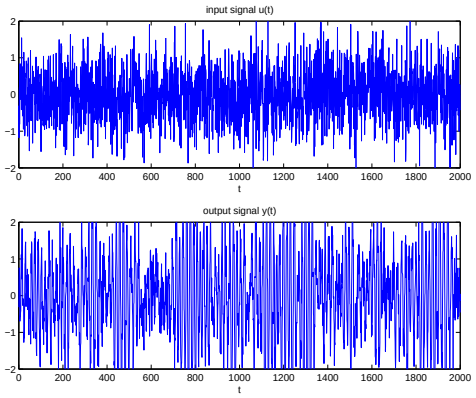
$G(z,\hat{\theta}_N)$ (blue) is compared with $G(z,\theta_0)$ (red):



Due to the lack of excitation, there are multiple θ^* which minimize $\bar{E}\epsilon^2(t,\theta)$ and the identified $\hat{\theta}_N$ is a consistent estimate of one of these θ^* ($\neq \theta_0$)

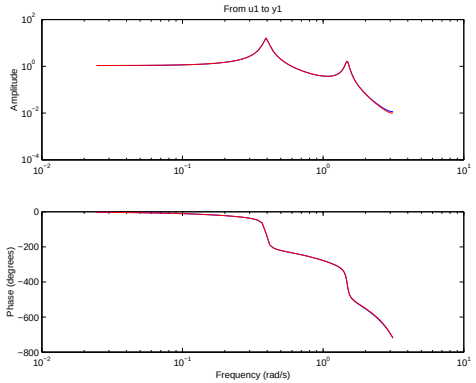
Second experiment on $\mathcal{S}$

we have applied a white noise $u(t)$ (u p.e. of order ∞) to $\mathcal{S}$ and collected $N = 2000$ IO data:



Using the 2000 recorded data, we have identified $\hat{\theta}_N$

$G(z, \hat{\theta}_N)$ (blue) is compared with $G(z, \theta_0)$ (red):



Since the signal u is p.e. of order ≥ 6, θ^* is unique and the identified $\hat{\theta}_N$ is a consistent estimate of this $\theta^* = \theta_0$

7 Statistical properties of $\hat{\theta}_N$ when $\mathcal{S} \in \mathcal{M}$

Due to the stochastic noise $v(t)$ corrupting the data Z^N, the identified parameter vector $\hat{\theta}_N$ is a random variable i.e.

the value of $\hat{\theta}_N$ is different at each experiment

When $\mathcal{S} \in \mathcal{M}$, the identified parameter vector $\hat{\theta}_N$ has the following property:

- $\hat{\theta}_N \sim \mathcal{N}(\theta_0, P_\theta)$
- $\hat{\theta}_N \to \theta_0$ with probability 1 when $N \to \infty$ (i.e. $P_\theta \to 0$ when $N \to \infty$).

Note: the first property is in fact $\hat{\theta}_N \sim As\mathcal{N}(\theta_0, P_\theta)$

7.1 Normal distribution of the identified parameter vector $\hat{\theta}_N$

Consider an identification experiment on $\mathcal{S}$ achieved using an input signal $u(t)$ and a number N of data

The parameter vector $\hat{\theta}_N$ identified in such an experiment is the realization of a normal distribution:

$$\hat{\theta}_N \sim \mathcal{N}(\theta_0, P_\theta)$$

$$\begin{aligned} P_\theta &\triangleq E\left((\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T\right) \\ &= \frac{\sigma_e^2}{N}\left(\bar{E}\psi(t,\theta_0)\psi^T(t,\theta_0)\right)^{-1} \end{aligned}$$

with $\psi(t,\theta_0) = \frac{\partial \hat{y}(t,\theta)}{\partial \theta}\Big|_{\theta=\theta_0} = -\frac{\partial \varepsilon(t,\theta)}{\partial \theta}\Big|_{\theta=\theta_0}$

Interpretation of $\hat{\theta}_N \sim \mathcal{N}(\theta_0, P_\theta)$

Consider p different identification experiments on $\mathcal{S}$ which deliver p different estimates $\hat{\theta}_N^{(i)}$

$E\hat{\theta}_N = \theta_0$ means that

$$\lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p} \hat{\theta}_N^{(i)} = \theta_0$$

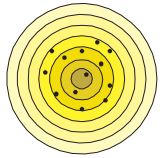
$\hat{\theta}_N$ unbiased estimate of θ_0

Interpretation of $\hat{\theta}_N \sim \mathcal{N}(\theta_0, P_\theta)$ (con't)

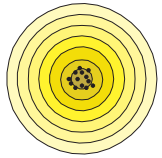
$$P_\theta \triangleq E\left((\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T\right)$$

the covariance matrix P_θ gives an idea of the standard deviation between $\hat{\theta}_N$ and θ_0 (see next slide)

Estimates $\hat{\theta}_N$ distributed as $\hat{\theta}_N \sim \mathcal{N}(\theta_0, P_\theta)$ with large P_θ



Estimates $\hat{\theta}_N$ distributed as $\hat{\theta}_N \sim \mathcal{N}(\theta_0, P_\theta)$ with small P_θ



Properties of the covariance matrix P_θ of $\hat{\theta}_N$

Property 1. P_θ is a function of the chosen input signal $u(t)$ and of the number N of data used for the identification.

Proof:

$$\epsilon(t,\theta) = \frac{G_0(z) - G(z,\theta)}{H(z,\theta)}u(t) + \frac{H_0}{H(z,\theta)}e(t) \Longrightarrow$$

$$\psi(t,\theta_0) = \left.\frac{-\partial\epsilon(t,\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \frac{\Lambda_G(z,\theta_0)}{H(z,\theta_0)}u(t) + \frac{\Lambda_H(z,\theta_0)}{H(z,\theta_0)}e(t)$$

with $\Lambda_G(z,\theta) = \frac{\partial G(z,\theta)}{\partial\theta}$ and $\Lambda_H(z,\theta) = \frac{\partial H(z,\theta)}{\partial\theta}$

$$\psi(t,\theta_0) = \frac{\Lambda_G(z,\theta_0)}{H(z,\theta_0)}u(t) + \frac{\Lambda_H(z,\theta_0)}{H(z,\theta_0)}e(t)$$

Now defining $\Gamma_G = \frac{\Lambda_G\Lambda_G^*}{HH^*}$ and $\Gamma_H = \frac{\Lambda_H\Lambda_H^*}{HH^*}$ and using Parseval theorem

$$P_\theta = \frac{\sigma_e^2}{N}\left(\bar{E}\psi(t,\theta_0)\psi^T(t,\theta_0)\right)^{-1} \Longrightarrow$$

$$P_\theta = \frac{\sigma_e^2}{N}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\Gamma_G(e^{j\omega},\theta_0)\ \Phi_u(\omega)\ + \Gamma_H(e^{j\omega},\theta_0)\ \sigma_e^2\ d\omega\right)^{-1}$$

$\Longrightarrow$ P_θ function of $u(t)$ and N

We can therefore influence the value of P_θ by appropriately choosing $u(t)$ and N

Property 2. The covariance matrix P_θ is a function of the unknown true system $\mathcal{S}$ via σ_e^2 and θ_0.

Property 3. A reliable estimate $\hat{P}_\theta$ of P_θ can nevertheless be deduced using the data and $\hat{\theta}_N$

$$\hat{P}_\theta = \frac{\hat{\sigma}_e^2}{N}\left(\frac{1}{N}\sum_{t=1}^{N}\psi(t,\hat{\theta}_N)\psi^T(t,\hat{\theta}_N)\right)^{-1}$$

with $\hat{\sigma}_e^2 = \frac{1}{N}\sum_{t=1}^{N}\epsilon(t,\hat{\theta}_N)^2$

7.2 Consistency property of the PEI estimate $\hat{\theta}_N$

$\hat{\theta}_N \to \theta_0$ with probability 1 when $N \to \infty$

$\equiv$

If we could collect $N = \infty$ data from $\mathcal{S}$, then the identified parameter vector $\hat{\theta}_{N\to\infty}$ would have the following distribution:

$$\hat{\theta}_{N\to\infty} \sim \mathcal{N}(\theta_0, P_\theta) \text{ with } P_\theta = 0$$

In other words, $\hat{\theta}_{N\to\infty}$ is a random variable whose realization is always equal to θ_0

Indeed:

$$P_\theta = \frac{\sigma_e^2}{N}\left(\bar{E}\psi(t,\theta_0)\psi^T(t,\theta_0)\right)^{-1}$$

and $N \to \infty \quad \Longrightarrow$

$$P_\theta \to 0$$

7.3 Proof of the statistical properties of $\hat{\theta}_N$ when $\mathcal{M}$ is FIR

$$\mathcal{S}: \quad G_0(z) = a_0 + b_0 z^{-1} \quad and \quad H_0(z) = 1$$

N input-output data have been collected from $\mathcal{S}$

Full-order FIR model structure:

$$G(z,\theta) = a + bz^{-1} \quad H(z,\theta) = 1$$

$$\theta = \begin{pmatrix} a \\ b \end{pmatrix}$$

Predictor:

$$\hat{y}(t,\theta) = \phi(t)^T\theta \quad with\ \phi^T(t) = \begin{pmatrix} u(t) & u(t-1) \end{pmatrix}$$

Note that the data $y(t)$ and $u(t)$ collected from $\mathcal{S}$ obey the following relation:

$$y(t) = \phi(t)^T \overbrace{\begin{pmatrix} a_0 \\ b_0 \end{pmatrix}}^{\theta_0} + e(t) \quad t = 1...N$$

The estimate $\hat{\theta}_N$ is obtained as follows:

$$\hat{\theta}_N = \underbrace{\left[\frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi^T(t)\right]^{-1}}_{R} \frac{1}{N}\sum_{t=1}^{N}\phi(t)y(t)$$

What is the relation between $\hat{\theta}_N$ and θ_0 ?

Replace $y(t)$ by its expression:

$$\hat{\theta}_N = R^{-1}\left(\frac{1}{N}\sum_{t=1}^{N}\phi(t)\overbrace{(\phi(t)^T\theta_0 + e(t))}^{=y(t)}\right)$$

$$\hat{\theta}_N = \theta_0 + \underbrace{R^{-1}\left(\frac{1}{N}\sum_{t=1}^{N}\phi(t)e(t)\right)}_{estimation\ error}$$

$$\Longrightarrow$$

$\hat{\theta}_N$ is a random variable and is (asymptotically) normally distributed

Indeed

- $e(t)$ is a random process and
- central limit theorem

What are the moments of this normal distribution ?

Mean:

$$E\hat{\theta}_N = \theta_0 + E\left(R^{-1}\left(\frac{1}{N}\sum_{t=1}^{N}\phi(t)e(t)\right)\right)$$

Since $\phi(t)$ and R are deterministic (not stochastic):

$$\begin{aligned} E\hat{\theta}_N &= \theta_0 + R^{-1}\left(\frac{1}{N}\sum_{t=1}^{N}\phi(t)\overbrace{Ee(t)}^{=0}\right) \\ &= \theta_0 \end{aligned}$$

Covariance matrix:

$$P_\theta \triangleq E\left((\hat{\theta}_N - \theta_0)(\hat{\theta}_N - \theta_0)^T\right)$$

$$\begin{aligned} P_\theta &= E\left(\frac{R^{-1}}{N}\left(\sum_{t=1}^{N}\phi(t)e(t)\right)\left(\sum_{s=1}^{N}e(s)\phi^T(s)\right)\frac{R^{-1}}{N}\right) \\ &= \frac{R^{-1}}{N}\left(\sum_{t=1}^{N}\sum_{s=1}^{N}\phi(t)\ E(e(t)e(s))\ \phi^T(s)\right)\frac{R^{-1}}{N} \\ &= \frac{R^{-1}}{N}\left(\sigma_e^2\sum_{t=1}^{N}\phi(t)\phi^T(t)\right)\frac{R^{-1}}{N} \\ &= \frac{\sigma_e^2}{N}R^{-1}RR^{-1} = \frac{\sigma_e^2}{N}R^{-1} \end{aligned}$$

The FIR case is a very particular case: only the normal distribution is asymptotic in N while $E\hat{\theta}_N = \theta_0$ and the covariance matrix are valid $\forall N$

Note that $P_\theta = \frac{\sigma_e^2}{N}R^{-1}$ converges when $N \to \infty$ to the asymptotic expression

$$\frac{\sigma_e^2}{N}\left(\bar{E}\psi(t,\theta_0)\psi^T(t,\theta_0)\right)^{-1}$$

since

$$\hat{y}(t,\theta) = \phi^T(t)\theta \Longrightarrow \psi(t,\theta) = \phi(t)\ \forall\theta$$

What happens when $N \to \infty$?

$$\hat{\theta}_{N\to\infty} = \theta_0 + \underbrace{\lim_{N\to\infty}\left(R^{-1}\frac{1}{N}\sum_{t=1}^{N}\begin{pmatrix} u(t)e(t) \\ u(t-1)e(t)\end{pmatrix}\right)}_{\text{random variable whose realisation is always 0}}$$

Parametric uncertainty region

$\hat{\theta}_N$ close to θ_0 if P_θ "small"

To determine how close, we can build an uncertainty region in the parameter space:

$$\hat{\theta}_N \sim \mathcal{N}(\theta_0, P_\theta) \iff$$

$$(\theta_0 - \hat{\theta}_N)^T\ P_\theta^{-1}\ (\theta_0 - \hat{\theta}_N)\ \sim\ \chi^2(k)$$

with k the dimension of $\hat{\theta}_N$

$$(\theta_0 - \hat{\theta}_N)^T \; P_\theta^{-1} \; (\theta_0 - \hat{\theta}_N) \;\sim\; \chi^2(k)$$

the unknown true parameter vector θ_0 lies therefore in the following ellipsoid U with probability, say, 95%

$$U = \left\{ \theta \in \mathbf{R}^k \mid (\theta - \hat{\theta}_N)^T \; P_\theta^{-1} \; (\theta - \hat{\theta}_N) \;\leq \alpha \right\}$$

with α such that $Pr(\chi^2(k) < \alpha) = 0.95$.

$$U = \left\{ \theta \in \mathbf{R}^k \mid (\theta - \hat{\theta}_N)^T \; P_\theta^{-1} \; (\theta - \hat{\theta}_N) \;\leq \alpha \right\}$$

The uncertainty ellipsoid U is centered at the identified parameter vector $\hat{\theta}_N$ and shaped by its covariance matrix P_θ

The largest P_θ, the largest the ellipsoid and thus the largest the uncertainty

Remark: $G(z, \theta_0)$ lies with the same probability in

$$\mathcal{D} = \{ G(z, \theta) \mid \theta \in U \}$$

Example:

$$\mathcal{S}: \quad y(t) = \frac{0.7 z^{-1}}{1 + 0.3 z^{-1}} u(t) + \frac{1}{1 + 0.3 z^{-1}} e(t)$$

$$\mathcal{M}: \quad G(z, \theta) = \frac{b z^{-1}}{1 + a z^{-1}} \quad H(z, \theta) = \frac{1}{1 + a z^{-1}} \quad \theta = \begin{pmatrix} a \\ b \end{pmatrix}$$

we have applied a sequence $u(t)$ of length $N = 1000$ to $\mathcal{S}$ and we have measured the corresponding $y(t)$.

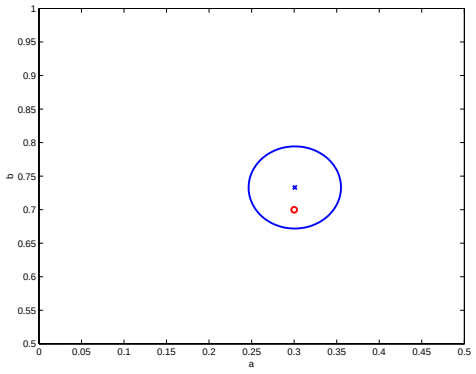
Using these data, we have computed the estimate $\hat{\theta}_N$ of $\theta_0 = \begin{pmatrix} 0.3 \;,\; 0.7 \end{pmatrix}^T$ along with its (estimated) covariance matrix P_θ:

$$\hat{\theta}_N = \begin{pmatrix} 0.301 \\ 0.733 \end{pmatrix} \qquad P_\theta = 10^{-3} \begin{pmatrix} 0.4922 & 0.0017 \\ 0.0017 & 0.6264 \end{pmatrix}$$

The 95% uncertainty region U can then be constructed

$$U = \left\{ \theta \in \mathbf{R}^k \mid (\theta - \hat{\theta}_N)^T \; P_\theta^{-1} \; (\theta - \hat{\theta}_N) \;\leq 5.99 \right\}$$

The estimate $\hat{\theta}_N$ (blue cross) along with its uncertainty ellipsoid U in the parameter space



The in practice unknown θ_0 is represented by the red circle and lies in U as expected

8 Statistical distribution of the identified model when $\mathcal{S} \in \mathcal{M}$

the identified parameter vector $\hat{\theta}_N$ is a random variable distributed as $\hat{\theta}_N \sim As\mathcal{N}(\theta_0, P_\theta) \Longrightarrow$

the identified models $G(z, \hat{\theta}_N)$ (and $H(z, \hat{\theta}_N)$) are also random variables:

- $G(z, \hat{\theta}_N)$ is an (asymptotically) unbiased estimate of $G(z, \theta_0)$
- the variance of $G(z, \hat{\theta}_N)$ is defined in the frequency domain as:

$$cov(G(e^{j\omega}, \hat{\theta}_N)) \triangleq E\left(|G(e^{j\omega}, \hat{\theta}_N) - G(e^{j\omega}, \theta_0)|^2\right)$$

$cov(G(e^{j\omega}, \hat{\theta}_N))$ can be expressed as a function of P_θ:

$$cov(G(e^{j\omega}, \hat{\theta}_N)) = \Lambda_G(e^{j\omega}, \theta_0)\ P_\theta\ \Lambda_G^*(e^{j\omega}, \theta_0)$$

with $\Lambda_G^T(z, \theta) = \frac{\partial G(z,\theta)}{\partial \theta}$

(obtained using a first order approximation and the assumption that N is large enough)

Properties of $cov(G(e^{j\omega}, \hat{\theta}_N))$

Property 1. $cov(G(e^{j\omega}, \hat{\theta}_N))$ is a function of the chosen $u(t)$ and of the number N of data used for the identification.

direct consequence of the fact that P_θ is a function of these quantities

More speaking relation between the choice of $u(t)$ and of N and $cov(G(e^{j\omega}, \hat{\theta}_N))$

Obtained by assuming that the MacMillan degree n of the model $G(z, \theta)$ in $\mathcal{M} \to \infty$

$$cov\left(G(e^{j\omega}, \hat{\theta}_N)\right) \approx \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

Property 2. $cov(G(e^{j\omega}, \hat{\theta}_N))$ **is a function of the unknown** $\mathcal{S}$

Property 3. **An estimate of** $cov(G(e^{j\omega}, \hat{\theta}_N))$ **can nevertheless be computed using the data and** $\hat{\theta}_N$

$$cov(G(e^{j\omega}, \hat{\theta}_N)) \approx \Lambda_G^*(e^{j\omega}, \hat{\theta}_N)\ \hat{P}_\theta\ \Lambda_G(e^{j\omega}, \hat{\theta}_N)$$

Comparison with non-parametric identification:

- $cov(G(e^{j\omega}, \hat{\theta}_N)) \to 0$ **when** $N \to \infty$ **(even for non-periodic signal)**
- **the modeling error at** ω_1 **is correlated to the error at** ω_2 **due to the parametrization**

9 Validation of the identified model when $\mathcal{S} \in \mathcal{M}$

We have identified a model $G(z, \hat{\theta}_N)$ in $\mathcal{M}$ using Z^N and we have verified that $\mathcal{S} \in \mathcal{M}$ (see later).

Important question: Is $G(z, \hat{\theta}_N)$ close to $G(z, \theta_0)$?

Validation using $cov(G(e^{j\omega}, \hat{\theta}_N)$

$$cov(G(e^{j\omega}, \hat{\theta}_N)) \triangleq E\left(|G(e^{j\omega}, \hat{\theta}_N) - G(e^{j\omega}, \theta_0)|^2\right)$$

Consequently, at each frequency ω:

the modeling error $|G(e^{j\omega}, \theta_0) - G(e^{j\omega}, \hat{\theta}_N)|$ is very likely to be small w.r.t. $|G(e^{j\omega}, \hat{\theta}_N)|$

if

the standard deviation $\sqrt{cov(G(e^{j\omega}, \hat{\theta}_N)}$ of $G(e^{j\omega}, \hat{\theta}_N)$ is small w.r.t. $|G(e^{j\omega}, \hat{\theta}_N)|$

More precizely, since $G(z, \hat{\theta}_N)$ is normally distributed, we have at each frequency ω that

$$|G(e^{j\omega}, \theta_0) - G(e^{j\omega}, \hat{\theta}_N)| < 1.96 \sqrt{cov(G(e^{j\omega}, \hat{\theta}_N)} \quad w.p.\ 95\%$$

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

What is a small standard deviation $\sqrt{cov(G(e^{j\omega}, \hat{\theta}_N)}$ (or a small modeling error) w.r.t. $|G(e^{j\omega}, \hat{\theta}_N)|$?

Highly dependent on the expected use for the model !!

For example, if we want to use the model for control, the modeling error (measured by $\sqrt{cov(G(e^{j\omega}, \hat{\theta}_N))}$) has to be much smaller around the cross-over frequency than at the other frequencies

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

What to do if the variance appears too large ?

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

A new identification experiment has then to be achieved in order to obtain a better model

For this purpose, we have to take care that the variance in this new identification is smaller

How can we reduce the variance of the identified model in a new identification ?

$$cov\left(G(e^{j\omega},\hat{\theta}_N)\right) \approx \frac{n}{N}\frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

Consequently, $cov(G(e^{j\omega},\hat{\theta}_N))$ can be reduced by

- **increasing the number of data N;**
- **or increasing the power spectrum $\Phi_u(\omega)$ of the input signal at the frequencies where $cov(G(e^{j\omega},\hat{\theta}_N))$ was too large**

Example

Let us consider the same flexible transmission system $\mathcal{S}$ (in the ARX form)

Let us consider a full order model structure $\mathcal{M}$ for $\mathcal{S}$

We want to use $G(z,\hat{\theta}_N)$ for control

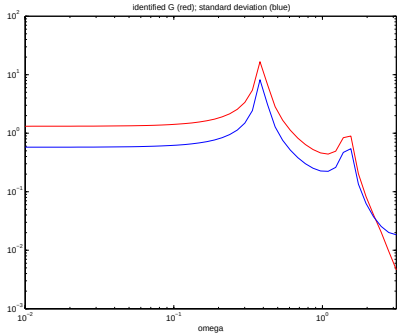
In this example, we need $\frac{\sqrt{cov(G(e^{j\omega},\hat{\theta}_N)}}{|G(e^{j\omega},\hat{\theta}_N)|} < 0.1\ \forall\omega \in [0\ 1]$

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 $G(z,\hat{\theta}_N)$ in $\mathcal{M}$

Validation of the identified model $G(z,\hat{\theta}_N)$:

we compare $\sqrt{cov(G(e^{j\omega},\hat{\theta}_N)}$ (blue) and $|G(z,\hat{\theta}_N)|$ (red):



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

Second identification experiment

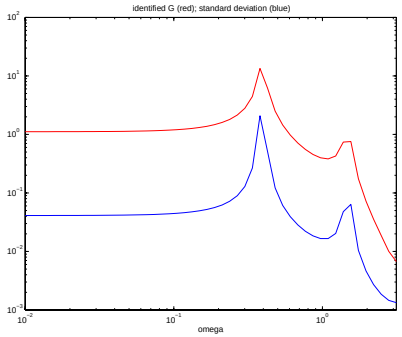
We want to reduce the variance of the identified model

Let us for this purpose increase the power of $u(t)$:

We apply a white noise input signal $u(t)$ of variance $\sigma_u^2 = 1$ to $\mathcal{S}$, collect $N = 2000$ IO data and identify a model $G(z,\hat{\theta}_N)$ in $\mathcal{M}$

Validation of the identified model $G(z, \hat{\theta}_N)$:

we compare $\sqrt{cov(G(e^{j\omega}, \hat{\theta}_N)}$ (blue) and $|G(z, \hat{\theta}_N)|$ (red):



$\sqrt{cov(G(e^{j\omega}, \hat{\theta}_N)}$ is better, but still too large at the 1st peak for our control purpose!!!

Third identification experiment

We want to reduce the variance of the identified model further around the 1st peak

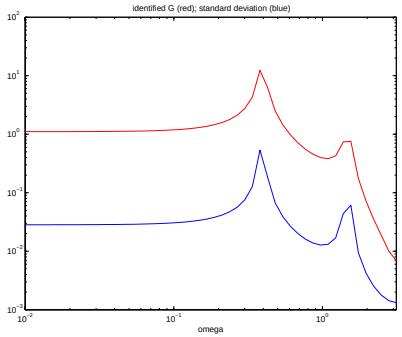
Let us for this purpose increase the power of $u(t)$ around this first peak:

$u(t) =$ 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 $G(z, \hat{\theta}_N)$ in $\mathcal{M}$

Validation of the identified model $G(z, \hat{\theta}_N)$:

we compare $\sqrt{cov(G(e^{j\omega}, \hat{\theta}_N)}$ (blue) and $|G(z, \hat{\theta}_N)|$ (red):



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

Final note:

Similar analysis can be made for $H(e^{j\omega}, \hat{\theta}_N)$ using $cov(H(e^{j\omega}, \hat{\theta}_N))$

$cov(H(e^{j\omega}, \hat{\theta}_N))$ can be deduced using a similar reasoning as for $cov(G(e^{j\omega}, \hat{\theta}_N))$

10 A special case of undermodelling

10.1 Identification in a model structure $\mathcal{M}$ which does not contain $\mathcal{S}$: $\mathcal{S} \notin \mathcal{M}$

$\mathcal{S} \notin \mathcal{M} \iff$ there does not exist a θ_0 such that

$$G(z,\theta_0) = G_0(z) \quad \text{and} \quad H(z,\theta_0) = H_0(z)$$

Consider a model structure $\mathcal{M} = \{G(z,\theta)\ ;\ H(z,\theta)\}$ such that $\mathcal{S} \notin \mathcal{M}$ and an input signal $u(t)$ sufficiently exciting of order $\geq n_g$

Define, as before, the ideal identification criterion:

$$\theta^* = \arg\min_{\theta} \bar{E}\epsilon^2(t,\theta)$$

and the estimate $\hat{\theta}_N$ of θ^* :

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

Statistical properties of $\hat{\theta}_N$ w.r.t. θ^*

- $\hat{\theta}_N \to \theta^*$ w.p. 1 when $N \to \infty$
- $\hat{\theta}_N \sim As\mathcal{N}(\theta^*, P_\theta)$ (P_θ having a more complicate expression than when $\mathcal{S} \in \mathcal{M}$)

Since $\mathcal{S} \notin \mathcal{M}$, we have in general:

$$G(z,\theta^*) \neq G_0(z) \quad and \quad H(z,\theta^*) \neq H_0(z)$$

One exception though:

$\mathcal{S} \notin \mathcal{M}$ with $G_0 \in \mathcal{G}$ and $\mathcal{M}$ OE, BJ or FIR

10.2 Special case of undermodelling: $\mathcal{S} \notin \mathcal{M}$ with $G_0 \in \mathcal{G}$

The model structure $\mathcal{M}$ used for identification purpose is such that

$$\exists \theta_0 \ \ such\ that \ \ G(z,\theta_0) = G_0(z) \ \ but \ \ H(z,\theta_0) \neq H_0(z)$$

What can be said about θ^* in this special case ?

To answer this question, we distinguish two classes of model structures $\mathcal{M}$:

- $\mathcal{M}$ with no common parameters in $G(\theta)$ and $H(\theta)$ (i.e. OE, BJ, FIR)

$$\theta = \begin{pmatrix} \eta \\ \zeta \end{pmatrix} \quad G(\theta) = G(\eta) \quad H(\theta) = H(\zeta)$$

- $\mathcal{M}$ with common parameters in $G(\theta)$ and $H(\theta)$ (i.e. ARX, ARMAX)

Result:

True system $\mathcal{S}$: $y = G_0 u(t) + H_0 e(t)$

Chosen model structure $\mathcal{M} = \{ G(z,\theta), H(z,\theta) \}$ such that $\exists \theta_0$ with $G(z,\theta_0) = G_0(z)$ but $H(z,\theta_0) \neq H_0(z)$.

- if $\mathcal{M}$ is OE, BJ or FIR, then

$$\theta^* = \begin{pmatrix} \eta^* \\ \zeta^* \end{pmatrix} \quad G(z,\eta^*) = G_0 \quad H(z,\zeta^*) \neq H_0$$

- if $\mathcal{M}$ is ARX or ARMAX, then

$$G(z,\theta^*) \neq \overbrace{G(z,\theta_0)}^{G_0} \quad H(z,\theta^*) \neq H_0$$

Example

$$y(t) = \frac{z^{-3}\left(0.103 + 0.181 z^{-1}\right)}{1 - 1.991 z^{-1} + 2.203 z^{-2} - 1.841 z^{-3} + 0.894 z^{-4}} u(t) + v(t)$$

with $v(t) = H_0(z)e(t)$; $H_0(z)$ very complicate i.e. $\mathcal{S}$ is not ARX, not OE !!!

We have applied a powerful white noise input signal ($\sigma_u^2 = 5$) to $\mathcal{S}$ and collected a large number of IO data ($N = 5000$) $\Longrightarrow$ small variance $\Longrightarrow \hat{\theta}_N \approx \theta^*$

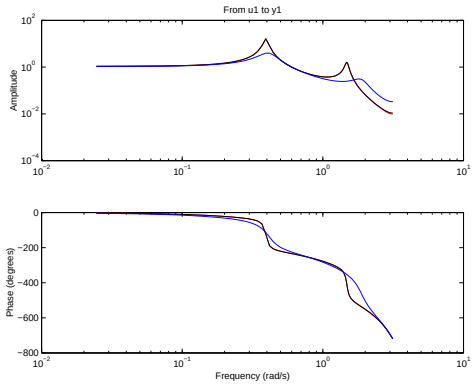
Using these IO data, we have identified a model in two model structures such that $\mathcal{S} \notin \mathcal{M}$ with $G_0 \in \mathcal{G}$:

$$\mathcal{M}_{arx} = ARX(n_a = 4, n_b = 2, n_k = 3)$$

$$\mathcal{M}_{oe} = OE(n_b = 2, n_f = 4, n_k = 3)$$

Let us denote $G(z, \hat{\theta}_N^{arx})$ and $G(z, \hat{\theta}_N^{oe})$, the models identified in $\mathcal{M}_{arx}$ and $\mathcal{M}_{oe}$, respectively.

Bode plots of $G(z, \hat{\theta}_N^{arx})$ (blue) and $G(z, \hat{\theta}_N^{oe})$ (black) and $G_0(z)$ (red)



As expected, we obtain $G(z, \hat{\theta}_N^{oe}) \approx G_0(z)$ and $G(z, \hat{\theta}_N^{arx})$ very different from G_0

11 Choice and validation of model order and structure

Until now, we have posed assumptions on the property of the model structure $\mathcal{M}$ w.r.t. $\mathcal{S}$:

- $\mathcal{S} \in \mathcal{M}$
- $\mathcal{S} \notin \mathcal{M}$ with $G_0 \in \mathcal{G}$
- $\mathcal{S} \notin \mathcal{M}$ with $G_0 \notin \mathcal{G}$

How can we verify these assumptions ?

a solution: model structure validation

11.1 Model structure validation: an a-posteriori verification

Assume that we have identified a parameter vector $\hat{\theta}_N$ in a model structure $\mathcal{M} = \{ G(\theta), H(\theta) \}$ with N data Z^N collected on the true system $\mathcal{S}$: $y(t) = G_0 u(t) + H_0 e(t)$.

Model structure validation: based on $\hat{\theta}_N$ and Z^N, determine if the chosen model structure $\mathcal{M}$ is such that:

- $\mathcal{S} \in \mathcal{M}$ or
- $\mathcal{S} \notin \mathcal{M}$ with $G_0 \in \mathcal{G}$
- $\mathcal{S} \notin \mathcal{M}$ with $G_0 \notin \mathcal{G}$

11.2 Model structure validation in the asymptotic case ($N \to \infty$)

The identified parameter vector is then θ^*

Model structure validation is performed by considering $R_\epsilon(\tau)$ and $R_{\epsilon u}(\tau)$ of $\epsilon(t, \theta^*)$:

$$\epsilon(t, \theta^*) = H(\theta^*)^{-1}(y(t) - G(\theta^*)u(t))$$

Due to the fact that

$$\epsilon(t, \theta^*) = \frac{G_0 - G(\theta^*)}{H(\theta^*)} u(t) + \frac{H_0}{H(\theta^*)} e(t),$$

three situations can occur for these quantities $R_\epsilon(\tau)$ and $R_{\epsilon u}(\tau)$

Situation A

We observe:

$$R_\epsilon(\tau) = \sigma_e^2 \delta(\tau) = \begin{cases} \sigma_e^2 & \text{for } \tau = 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$R_{\epsilon u}(\tau) = 0 \quad \forall\ \tau$$

This situation occurs when

$$\begin{aligned} \epsilon(t, \theta^*) &= \frac{G_0 - G(\theta^*)}{H(\theta^*)} u(t) + \frac{H_0}{H(\theta^*)} e(t) \\ &= 0 \times u(t) + e(t) \end{aligned}$$

$$\iff G(\theta^*) = G_0 \quad \text{and} \quad H(\theta^*) = H_0$$

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

Situation B

We observe:

$$R_\epsilon(\tau) \neq \sigma_e^2 \delta(\tau)$$
$$R_{\epsilon u}(\tau) = 0 \quad \forall\, \tau$$

This situation occurs when

$$\epsilon(t,\theta^*) = \frac{G_0 - G(\theta^*)}{H(\theta^*)} u(t) + \frac{H_0}{H(\theta^*)} e(t)$$
$$= 0 \times u(t) + \frac{\overbrace{H_0}^{\neq 1}}{H(\theta^*)} e(t)$$
$$\iff G(\theta^*) = G_0 \text{ and } H(\theta^*) \neq H_0$$
$$\iff \mathcal{S} \notin \mathcal{M} \text{ with } G_0 \in \mathcal{G} \text{ for } \mathcal{M} \text{ OE, BJ or FIR}$$

Situation C

We observe:

$$R_\epsilon(\tau) \neq \sigma_e^2 \delta(\tau)$$
$$\exists \tau \;\; s.t. \;\; R_{\epsilon u}(\tau) \neq 0$$

This situation occurs when

$$\epsilon(t,\theta^*) = \frac{\overbrace{G_0 - G(\theta^*)}^{\neq 0}}{H(\theta^*)} u(t) + \frac{H_0}{H(\theta^*)} e(t)$$
$$\iff G(\theta^*) \neq G_0$$
$$\iff \begin{cases} \text{either } \mathcal{S} \notin \mathcal{M} \text{ with } G_0 \in \mathcal{G} \text{ for } \mathcal{M} \text{ ARX or ARMAX} \\ \text{or } \mathcal{S} \notin \mathcal{M} \text{ with } G_0 \notin \mathcal{G} \end{cases}$$

Conclusions for the asymptotic case:

1) $\mathcal{M}$ is chosen as OE, FIR or BJ:

Situations A, B and C can occur for $R_\epsilon(\tau)$ and $R_{\epsilon u}(\tau)$

By determining in which situations we are, we verify whether the identification of θ^* has been performed in a $\mathcal{M}$ such that

- $\mathcal{S} \in \mathcal{M}$ (situation A)
- $\mathcal{S} \notin \mathcal{M}$ with $G_0 \in \mathcal{G}$ (situation B)
- or $\mathcal{S} \notin \mathcal{M}$ with $G_0 \notin \mathcal{G}$ (situation C)

2) $\mathcal{M}$ is chosen as ARX or ARMAX:

Situations A and C can occur for $R_\epsilon(\tau)$ and $R_{\epsilon u}(\tau)$

By determining in which situations we are, we verify whether the identification of θ^* has been performed in a $\mathcal{M}$ such that

- $\mathcal{S} \in \mathcal{M}$ (situation A)
- $\mathcal{S} \notin \mathcal{M}$ (situation C)

No distinction can be made between $G_0 \in \mathcal{G}$ and $G_0 \notin \mathcal{G}$

11.3 Model structure validation in the practical case $N < \infty$

The identified parameter vector is $\hat{\theta}_N$ which is an unbiased estimate of θ^*

Model structure validation is performed by considering $\hat{R}_\epsilon^N(\tau)$ and $\hat{R}_{\epsilon u}^N(\tau)$ of $\epsilon(t, \hat{\theta}_N)$:

$$\hat{R}_{\epsilon u}^N(\tau) = \frac{1}{N}\sum_{t=1}^{N-\tau} \epsilon(t+\tau, \hat{\theta}_N)u(t)$$
$$\hat{R}_{\epsilon}^N(\tau) = \frac{1}{N}\sum_{t=1}^{N-\tau} \epsilon(t+\tau, \hat{\theta}_N)\epsilon(t, \hat{\theta}_N)$$

and by considering 99%-confidence regions for these estimates

What do these 99%-confidence regions represent ?

$\hat{R}_\epsilon^N(\tau)$ lies in its confidence region $\forall\tau \overset{\text{w.p.}}{\Longrightarrow} R_\epsilon(\tau) = \sigma_e^2\delta(\tau)$

$\hat{R}_{\epsilon u}^N(\tau)$ lies in its confidence region $\forall\tau \overset{\text{w.p.}}{\Longrightarrow} R_{\epsilon u}(\tau) = 0\ \forall\tau$

To construct these confidence regions, we use the following result:

if $R_\epsilon(\tau) = \sigma_e^2\delta(\tau)$, then $\sqrt{N}\frac{\hat{R}_\epsilon^N(\tau)}{\hat{R}_\epsilon^N(0)} \sim As\mathcal{N}(0,1)$.

if $R_{\epsilon u}(\tau) = 0 \ \ \forall\ \tau$, then $\sqrt{N}\hat{R}_{\epsilon u}^N(\tau) \sim As\mathcal{N}(0,P)$ with an estimable P.

Based on the results of the asymptotic case, we can therefore deduce:

1) when $\mathcal{M}$ is OE, FIR, or BJ

both $\hat{R}_\epsilon^N(\tau)$ and $\hat{R}_{\epsilon u}^N(\tau)$ are in their confidence regions $\forall\tau$ $\overset{\text{w.p.}}{\Longrightarrow}$ $\mathcal{S} \in \mathcal{M}$

$\hat{R}_{\epsilon u}^N(\tau)$ is in its confidence regions $\forall\tau$ while $\hat{R}_\epsilon^N(\tau)$ is not completely in its confidence region $\overset{\text{w.p.}}{\Longrightarrow}$ $\mathcal{S} \notin \mathcal{M}$ with $G_0 \in \mathcal{G}$

both $\hat{R}_\epsilon^N(\tau)$ and $\hat{R}_{\epsilon u}^N(\tau)$ are not completely in their confidence regions $\overset{\text{w.p.}}{\Longrightarrow}$ $\mathcal{S} \notin \mathcal{M}$ with $G_0 \notin \mathcal{G}$

2) when $\mathcal{M}$ is ARX or ARMAX

both $\hat{R}_{\epsilon}^{N}(\tau)$ and $\hat{R}_{\epsilon u}^{N}(\tau)$ are in their confidence regions $\forall \tau$ $\overset{\text{w.p.}}{\Longrightarrow}$ $\mathcal{S} \in \mathcal{M}$

other cases $\overset{\text{w.p.}}{\Longrightarrow}$ $\mathcal{S} \notin \mathcal{M}$

No distinction can be made between $G_0 \in \mathcal{G}$ and $G_0 \notin \mathcal{G}$

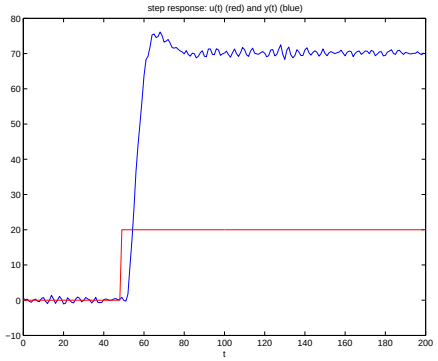
11.4 Example of how we can find a $\mathcal{M}$ s.t. $\mathcal{S} \in \mathcal{M}$

Let us consider an unknown true system $\mathcal{S}$

We would like to determine a model set $\mathcal{M}$ which contains $\mathcal{S}$

First analysis of the system

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



From this behaviour, we can conclude that G_0 has a limited order and from a detailed observation, we see that the delay is $n_k = 3$

Collection of the data for the identification and determination of $\mathcal{M}$

We have applied a white noise input signal to $\mathcal{S}$ and collected $N = 5000$ input-output data $\Longrightarrow Z^N$

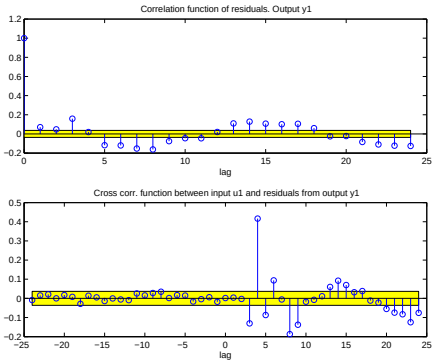
Based on the first analysis of $\mathcal{S}$, first choice for $\mathcal{M}$:

$$\mathcal{M} = BJ(n_b = 2, n_c = 2, n_d = 2, n_f = 2, n_k = 3)$$

We can identify a parameter vector $\hat{\theta}_N$ in this $\mathcal{M}$ using Z^N

Does this $\mathcal{M}$ contain the true system $\mathcal{S}$?

Let us perform the model structure validation (Matlab function: resid)



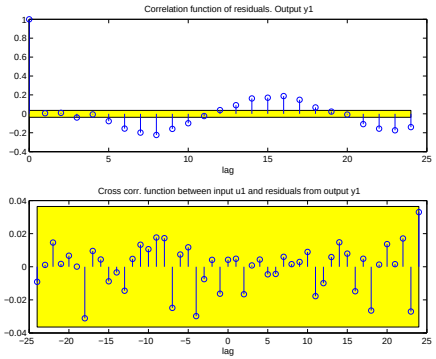
$\stackrel{\text{w.p.}}{\Longrightarrow} \mathcal{S} \notin \mathcal{M}$ **with** $G_0 \notin \mathcal{G}$

Let us increase the order for $G(z,\theta)$ and $H(z,\theta)$

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

and identify $\hat{\theta}_N$ in this new model structure using the same data Z^N

Let us perform the model structure validation of this new $\mathcal{M}$:



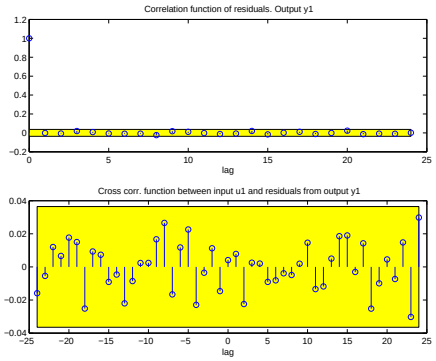
$\stackrel{\text{w.p.}}{\Longrightarrow} \mathcal{S} \notin \mathcal{M}$ **with** $G_0 \in \mathcal{G}$

A third order $H(z,\theta)$ is thus not sufficient to describe $H_0(z)$. Let us try:

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

and identify $\hat{\theta}_N$ in this new model structure using the data Z^N

Let us perform the model structure validation of this new $\mathcal{M}$:



$\stackrel{\text{w.p.}}{\Longrightarrow} \mathcal{S} \in \mathcal{M}$

By a simple iteration, we can find a model set $\mathcal{M}$ that has the property $\mathcal{S} \in \mathcal{M}$

Note: the used $\mathcal{S}$ was indeed BJ(3,4,4,3,3) !!

11.5 Final remarks.

Model structure validation validates the hypothesis $\mathcal{S} \in \mathcal{M}$ based on the available data

Other data can be used for the validation than for the identification

Model structure validation is often called model validation

However

a successful model structure validation does not necessarily imply that $G(z, \hat{\theta}_N)$ and $H(z, \hat{\theta}_N)$ are close estimates of $G_0(z) = G(z, \theta_0)$ and $H_0(z) = H(z, \theta_0)$ (variance can be still large !!!)

12 A typical procedure to identify a reliable full-order model

For some type of systems, a reasonable objective can be to identify reliable full-order models $G(z, \hat{\theta}_N)$ and $H(z, \hat{\theta}_N)$ of G_0 and H_0

To reach this objective:

Model structure validation allows to determine a model set $\mathcal{M}$ such that $\mathcal{S} \in \mathcal{M}$

and $\sqrt{cov(G(e^{j\omega}, \hat{\theta}_N))}$ allows one to verify whether $G(z, \hat{\theta}_N)$ is close to G_0 (and eventually $\sqrt{cov(H(e^{j\omega}, \hat{\theta}_N))}$ for $H(z, \hat{\theta}_N)$)

$\Longrightarrow$ **Typical iterative procedure**

1. choose the input signal and collect Z^N
2. choose a model structure $\mathcal{M}$
3. identification of the models $G(z, \hat{\theta}_N)$ and $H(z, \hat{\theta}_N)$
4. Verify if $\mathcal{S} \in \mathcal{M}$. If it is the case, go to item 5. If not, go to item 2 and choose another model structure $\mathcal{M}$
5. Verify if $\sqrt{cov(G(e^{j\omega}, \hat{\theta}_N))}$ (and eventually $\sqrt{cov(H(e^{j\omega}, \hat{\theta}_N))}$) are small. If not, go back to item 1. If yes, stop

Possible additional tests for item 5:

- simulation of the identified model
- observation of the poles and zeros of the identified models
- comparison of the frequency response of the identified models with the ETFE (see later) and/or with the physical equations.

13 Identification in a low order model structure

Some real-life systems have a very large order (e.g. chemical and industrial plants)

For such plants, identifying a reliable full-order model is:

- not a good idea since $cov(G(e^{j\omega}, \hat{\theta}_N))$ will be typically very large

$$cov(G(e^{j\omega}, \hat{\theta}_N)) \approx \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

with n large and N, $\Phi_u(\omega)$ limited

- not necessary: for control, a low order model accurate in the frequencies around the cross-over frequency is sufficient

$$\Longrightarrow$$

For that type of $\mathcal{S}$,

- choose a reduced order $\mathcal{M}$ which is nevertheless sufficiently rich to be able to represent the behaviour of the system in the important frequency range
- perform the identification experiment in such a way that the identified model is a close estimate of $\mathcal{S}$ in the important frequency range

Considered problem: What is the influence of the experimental conditions (choice of $u(t)$, choice of N) on the approximation of $G_0(z)$ by $G(z, \hat{\theta}_N)$ when:

$$\mathcal{S}: \quad y(t) = G_0(z)u(t) + \overbrace{H_0(z)e(t)}^{v(t)}$$

and $\mathcal{M} = \{G(z,\theta)\ ;\ H(z,\theta) = 1\}$ is an OE model structure such that $\nexists \theta_0$ with $G(z,\theta_0) = G_0(z)$

We restrict thus attention to:

- to the approximation of G_0 by $G(z, \hat{\theta}_N)$
- to Output Error (OE) model structure $\mathcal{M}$ (reason: easier analysis)

Reminder from before ....

$\hat{\theta}_N$ can be computed as in the case $\mathcal{S} \in \mathcal{M}$

$\hat{\theta}_N$ is a random variable due to the stochastic disturbance $v(t)$ corrupting the data

$\hat{\theta}_N$ is distributed as $\mathcal{N}(\theta^*, P_\theta)$ where θ^* is the solution of the ideal identification criterion

P_θ can not be determined analytically, but $P_\theta \to 0$ when $N \to \infty \Longrightarrow \hat{\theta}_N \to \theta^*$ w.p. 1 when $N \to \infty$

$\nexists \theta_0$ with $G(z, \theta_0) = G_0(z) \Longrightarrow G(z, \theta^*) \neq G_0(z)$

13.1 Modeling error when $\mathcal{S} \notin \mathcal{M}$ with $G_0 \notin \mathcal{G}$

the modeling error $G_0(z) - G(z, \hat{\theta}_N)$ is decomposed into two contributions:

$$G_0(z) - G(z, \hat{\theta}_N) = (G_0(z) - G(z, \theta^*)) + \Big(G(z, \theta^*) - G(z, \hat{\theta}_N)\Big)$$

Note: when $\mathcal{S} \in \mathcal{M}$, $G_0(z) - G(z, \theta^*) = 0$

the two contributions and their source:

$$G_0(z) - G(z, \hat{\theta}_N) = (G_0(z) - G(z, \theta^*)) + \Big(G(z, \theta^*) - G(z, \hat{\theta}_N)\Big)$$

- $G_0 - G(\theta^*)$ is called the **bias error** and is due to the fact that $\mathcal{S} \notin \mathcal{M}$ with $G_0 \notin \mathcal{G}$;
- $G(\theta^*) - G(\hat{\theta}_N)$ is called the **variance error** and is due to the fact that $N < \infty$

Considered problem (rephrased): what is the influence of the experimental conditions

- on the bias error
- on the variance error

13.3 shaping the bias error $G_0 - G(\theta^*)$

Recall we consider an OE model structure $\mathcal{M}$

13.3.1 a frequency domain expression of the bias error $G_0(e^{j\omega}) - G(e^{j\omega}, \theta^*)$

$$\theta^* = \arg\min_\theta \bar{V}(\theta)$$

and

$$\begin{aligned} \bar{V}(\theta) &= \bar{E}\epsilon(t,\theta)^2 \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_\epsilon(\omega,\theta)d\omega \end{aligned}$$

(Parseval; both expressions are equal to $R_\epsilon(0)$)

$\mathcal{M}$ = OE $\Longrightarrow$

$$\epsilon(t,\theta) = (G_0(z) - G(z,\theta))u(t) + v(t)$$

$\Longrightarrow$

$$\begin{aligned} \theta^* &= \arg\min_\theta \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_\epsilon(\omega,\theta)d\omega \\ &= \arg\min_\theta \frac{1}{2\pi}\int_{-\pi}^{\pi} |G_0(e^{j\omega}) - G(e^{j\omega},\theta)|^2\Phi_u(\omega) + \Phi_v(w)d\omega \end{aligned}$$

$$\theta^* = \arg\min_\theta \frac{1}{2\pi}\int_{-\pi}^{\pi} |G_0(e^{j\omega}) - G(e^{j\omega},\theta)|^2\Phi_u(\omega) + \Phi_v(w)d\omega$$

$\Longrightarrow$

$G(e^{j\omega}, \theta^*)$ is the model minimizing the integrated quadratic error $|G_0(e^{j\omega}) - G(e^{j\omega},\theta)|^2$ with weighting function $\Phi_u(\omega)$

$\Longrightarrow$

the bias will be the smallest at those ω's where $\Phi_u(\omega)$ is relatively the largest

Notes:

- the bias error is a function of the power spectrum $\Phi_u(\omega)$ of the input signal used for the identification
- the bias obtained with a signal $u(t)$ of spectrum $\Phi_u(\omega)$ is the same as the bias obtained with spectrum $\alpha\Phi_u(\omega)$ (α a scalar constant)
- the absolute level of power has thus no influence on the bias error, but influences the variance error

13.3.2 Another way to shape the bias error - off-line prefiltering

Given a filter $L(z)$ and the data $u(t)$ and $y(t)$ collected from $\mathcal{S}$

Filter $u(t)$ and $y(t)$ with L:

$$u_F(t) = L(z)u(t) \quad \text{and} \quad y_F(t) = L(z)y(t)$$

Result:

If you use the data $u_F(t)$ and $y_F(t)$ for the identification, the weighting function shaping the bias error is:

$$W(\omega) = \Phi_u(\omega)|L(e^{i\omega})|^2$$

Proof:

If we use the data $u_F(t)$ and $y_F(t)$ for the identification, the corresponding prediction error $\epsilon_F(t,\theta)$ is

$$\epsilon_F(t,\theta) = L(z)\epsilon(t,\theta)$$

where $\epsilon(t,\theta)$ is the prediction error if we would have used $u(t)$ and $y(t)$

Consequently,

$$\Phi_{\epsilon_F}(\omega,\theta) = |L(e^{i\omega})|^2 \cdot \Phi_\epsilon(\omega,\theta)$$

and therefore $W(\omega) = \Phi_u(\omega)|L(e^{i\omega})|^2$

13.4 shaping the variance error $G(\theta^*) - G(\hat{\theta}_N)$

Analysis more difficult than in the case $\mathcal{S} \in \mathcal{M}$

However we can nevertheless cautiously state that

- large $\Phi_u(\omega)$ around $\omega \Longrightarrow$ small variance error around ω
- large $N \Longrightarrow$ small variance error

13.5 Example

$$\mathcal{S}: \quad y(t) = G_0(z)u(t) + e(t)$$

with $G_0(z)$ 4th order with three delay

We have to use a given set of data Z^N $(N = 5000)$ for the identification where u is the sum of a white noise of variance 5 and three high-frequencies sinus of amplitude 10

Objective: Using the given data, identify a good model $G(z,\hat{\theta}_N)$ for $G_0(z)$ in the frequency range $[0\ 0.7]$ in the reduced order model structure:

$$\mathcal{M} = OE(n_b = 2, n_f = 2, n_k = 3)$$

Since Z^N is given, the only degree of freedom we have is to use a pre-filter $L(z)$ to shape the bias error

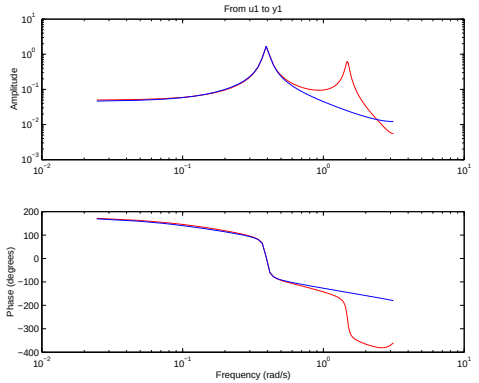
We want a small bias error in the frequency range $[0\ 0.7] \Longrightarrow$

choose $L(z)$ such that $|L(e^{j\omega})|^2\Phi_u(\omega)$ is relatively (much) larger in the frequency range $[0\ 0.7]$ than in $[0.7\ \pi]$

$\Longrightarrow$ $L(z)$ Butterworth low pass filter of order 7 and cut-off frequency $0.7rad/s$

We filter u and y collected from $\mathcal{S}$ by this L and we obtain filtered data with which we perform the identification in $\mathcal{M}$

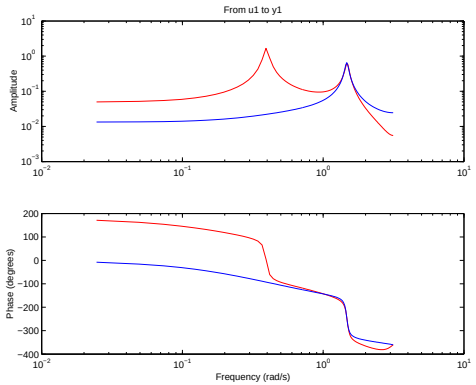
$G_0(z)$ (red) and $G(z,\hat{\theta}_N)$ (blue) identified with the filtered data



$\Longrightarrow$ $G(z,\hat{\theta}_N)$ is OK

What if we do not use a pre-filter L ?

$G_0(z)$ (red) and $G(z,\hat{\theta}_N)$ (blue) identified with the data in Z^N



$\Longrightarrow$ $G(z,\hat{\theta}_N)$ is KO

13.6 What about a Box Jenkins model structure

The weighting function $W(\omega)$ for the bias error $G_0(e^{j\omega}) - G(e^{j\omega},\theta^*)$ is then

$$W(\omega) = \frac{\Phi_u(\omega)|L(e^{i\omega})|^2}{|H(e^{j\omega},\theta^*)|^2}$$

the noise model $H(e^{j\omega},\theta^*)$ influences the bias error of the G-model !!

Part IV: Nonparametric Identification (ETFE)

General objective of ETFE

$$\mathcal{S}: \quad y(t) = G_0(z)u(t) + v(t)$$

We apply an input signal $u(t)$ to $\mathcal{S}$ and we collect the corresponding output for N time samples:

$$Z^N = \{\ y(t),\ u(t) \mid t = 0...(N-1)\ \}$$

Based on these N time-domain data, we want to estimate the frequency response $G_0(e^{j\omega})$ (amplitude and phase) of the true plant transfer function

Nonparametric identification is generally performed in order

- to have a first idea of $G_0(e^{j\omega})$
- to determine the frequency band of interest

Empirical Transfer Function Estimate (ETFE)

Time-Domain data $\longrightarrow$ Frequency-Domain data via (scaled) Fourier Transform

$$\{\ u(t) \mid t = 0...(N-1)\ \} \longleftrightarrow U_N(\omega) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} u(t)\ e^{-j\omega t}$$

$$\{\ y(t) \mid t = 0...(N-1)\ \} \longleftrightarrow Y_N(\omega) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y(t)\ e^{-j\omega t}$$

Estimate $\hat{G}(e^{j\omega})$ of $G_0(e^{j\omega})$

$$\hat{G}(e^{j\omega}) = |\hat{G}(e^{j\omega})|e^{j\angle\hat{G}(e^{j\omega})} = \frac{Y_N(\omega)}{U_N(\omega)}$$

$\hat{G}(e^{j\omega})$ can in theory be computed at each frequency $\omega \in [0\ \pi]$ for which $U_N(\omega) \neq 0$

Practical Aspects

All information contained in $\{\ u(t) \mid t = 0...(N-1)\ \}$ is contained in the elements of $U_N(\omega)$ at the $\frac{N}{2}$ frequencies $\omega_k = \frac{2\pi}{N}k$, $k = 0, 1, ...$ located in $[0\ \pi]$

$\hat{G}(e^{j\omega})$ is therefore only computed at those frequencies ω_k

Special attention should be given when $u(t)$ is a periodic signal of fundamental frequency ω_0

The Fourier transform $U_N(\omega)$ of such a signal is indeed only significant at the (active) harmonics of ω_0. $\hat{G}(e^{j\omega})$ will therefore only be computed at those harmonics.

Illustration

$$y(t) = \overbrace{\frac{z^{-3}\left(0.103 + 0.181z^{-1}\right)}{1 - 1.991z^{-1} + 2.203z^{-2} - 1.841z^{-3} + 0.894z^{-4}}}^{G_0(z)} u(t) + H_0 e(t)$$

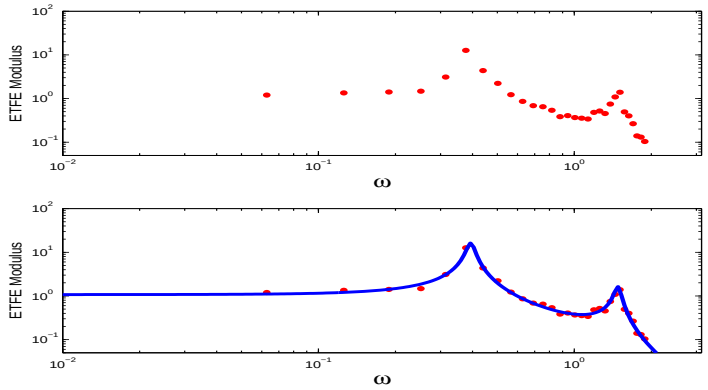
with $H_0 = 1/den(G_0)$ and $e(t)$ a white noise disturbance of variance $\sigma_e^2 = 0.1$

We collect $N = 10000$ data on this true system subsequently with two different input signals having the same $\mathcal{P}_u = 0.5 = 5\sigma_e^2$

Input signal 1: a multisine of fundamental frequency $\omega_0 = \frac{2\pi}{100} \approx 0.06$ (power=0.5)

$$u(t) = \frac{1}{\sqrt{30}} \sum_{k=1}^{30} sin(k\omega_0 t)$$

The ETFE is computed at the 30 harmonics of ω_0 present in $u(t)$

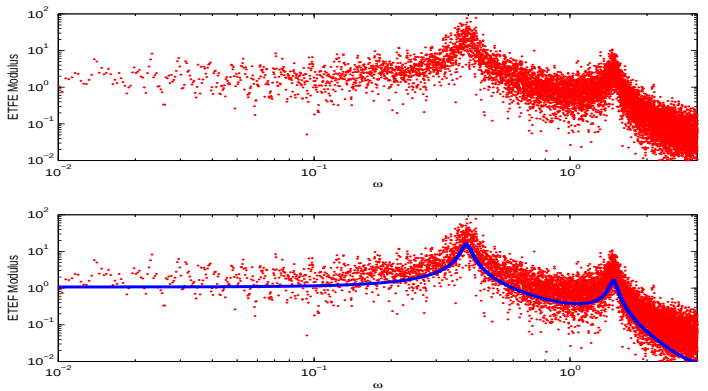


Above plot: the ETFE at the 30 harmonics of ω_0; Bottom plot: the same with the frequency response of $G_0(e^{j\omega})$ (blue)

We see that the ETFE is a good estimate of $G_0(e^{j\omega})$ at the harmonics of ω_0

Input signal 2: a white noise of variance 0.5

The ETFE is computed at all the $\frac{N}{2} = 5000$ frequencies ω_k



Above plot: the ETFE at ω_k; Bottom plot: the same with the frequency response of $G_0(e^{j\omega})$ (blue)

We see that the ETFE is an erratic and poor estimate of $G_0(e^{j\omega})$

How can we explain this?

For this purpose, we need to understand the statistical properties of the ETFE

Statistical properties of the ETFE

Due to the stochastic noise $v(t)$ corrupting the data Z^N, the ETFE $\hat{G}(e^{j\omega})$ is a random variable i.e.

the ETFE is different at each experiment

Moreover,

there is no (cor)relation between the estimate at the frequency ω_k and the other frequencies i.e. ω_{k-1}, ω_{k+1}, ...

At one frequency ω_k, the estimate $\hat{G}(e^{j\omega_k})$ is a random variable (asymptotically) distributed around $G_0(e^{j\omega_k})$

$$\Longrightarrow$$

the ETFE will be reliable if the variance of the estimates $\hat{G}(e^{j\omega_k})$ are small for all ω_k

Variance of the ETFE

the variance $cov(\hat{G}(e^{j\omega})) \triangleq E|\hat{G}(e^{j\omega}) - E\hat{G}(e^{j\omega})|^2$ is given by:

$$cov(\hat{G}(e^{j\omega})) = E\left(\frac{|V_N(e^{j\omega})|^2}{|U_N(e^{j\omega})|^2}\right)$$

with $V_N(\omega)$ defined as $Y_N(\omega)$ and $U_N(\omega)$

$cov(\hat{G}(e^{j\omega}))$ tends, for increasing values of N, to $\frac{\Phi_v(\omega)}{\Phi_u(\omega)}$

Explanation of the results in the illustration

Multisine: $u(t) = \frac{1}{\sqrt{30}} \sum_{k=1}^{30} sin(k\omega_0 t)$

The ETFE is only computed at the harmonics $\omega_k = k\ \omega_0$ $(k = 1...30)$ of ω_0.

Property of $|U_N|^2$ at the harmonics ω_k:

$$\left(|U_N(e^{j\omega_k})|^2\right) = \frac{NA_k^2}{4} = \frac{10000}{120}$$

since the amplitude A_k of each sine is $1/\sqrt{30}$ and $N = 10000$

What is the variance of the ETFE at the available frequencies ω_k ?

$E|U_N|^2 = |U_N|^2$ since $u(t)$ is deterministic

$\Longrightarrow$

$$cov(\hat{G}(e^{j\omega_k})) = \frac{E\left(|V_N(e^{j\omega_k})|^2\right)}{|U_N(e^{j\omega_k})|^2} \approx \frac{\Phi_v(\omega_k)}{|U_N(e^{j\omega_k})|^2} = \frac{120\Phi_v(\omega_k)}{10000}$$

Since $|U_N|^2$ is proportional to N and A_k^2, the variance is proportional to $\frac{1}{N}$ and $\frac{1}{A_k^2}$

$u(t)$ white noise of variance $\sigma_u^2 = 0.5$

The ETFE is computed at $\frac{N}{2} = 5000$ frequencies ω_k

Since N is large, the variance at the frequencies ω_k can be approximated by:

$$cov(\hat{G}(e^{j\omega_k})) \approx \frac{\Phi_v(\omega_k)}{\Phi_u(\omega_k)} = \frac{\Phi_v(\omega_k)}{\sigma_u^2} = \frac{\Phi_v(\omega_k)}{0.5} = 2\Phi_v(\omega_k)$$

Unlike for a multisine $u(t)$, the variance is not proportional to $\frac{1}{N}$; variance only proportional to $\frac{1}{\sigma_u^2}$

Multisine vs. stochastic signal

ETFE available at more frequencies for stochastic $u(t)$

For equal power, variance much smaller for multisine $u(t)$

Suppose $u(t)$ is not free to be chosen and is stochastic,

and that the power of $u(t)$ cannot be increased

How can we then get a relatively good estimate? How can we reduce the variance ?

Smoothing of ETFE through the use of windows

only really relevant when $u(t)$ is stochastic

Principle: reduction of the variance by averaging over neighbouring frequency points

Smoothing is motivated by:

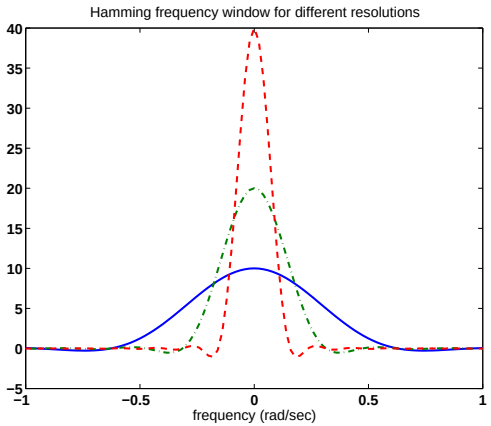
- ETFE estimates are independent for different ω_k's
- Averaging over a frequency area where G_0 is constant reduces the variance

The averaging can be performed as follows:

$$\hat{G}_{sm}(e^{j\omega}) = \frac{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega)\hat{G}(e^{i\xi})d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega)d\xi}$$

with $\hat{G}(e^{j\omega})$ the unsmoothed ETFE and $W_\gamma(\omega)$ a positive real-valued frequency-function (window)

A Hamming window is generally chosen for $W_\gamma(\omega)$



$W_\gamma(\omega)$ of Hamming window for $\gamma = 10$ (solid), $\gamma = 20$ (dash-dotted) and $\gamma = 40$ (dashed).

γ is measure for the width of the window.

The window is non zero in an interval $[-\Delta\omega,\ +\Delta\omega]$ around 0.

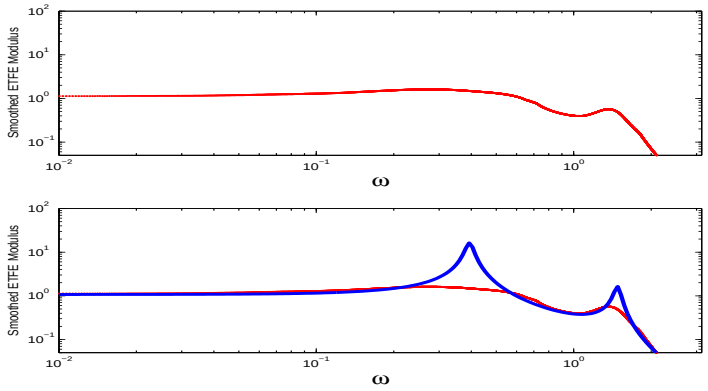
The larger γ, the smaller $\Delta\omega$.

$$\hat{G}_{sm}(e^{j\omega}) = \frac{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega)\hat{G}(e^{i\xi})d\xi}{\int_{-\pi}^{\pi} W_\gamma(\xi - \omega)d\xi}$$

$\hat{G}_{sm}(e^{j\omega_k})$ at a particular frequency ω_k is obtained by averaging $\hat{G}(e^{j\omega})$ in the interval $[\omega_k - \Delta\omega,\ \ \omega_k + \Delta\omega]$

- Window introduced **bias** in an attempt to reduce the **variance** (bias/variance trade-off)
- Choice of window dependent on expected smoothness of $G_0(e^{i\omega})$
- Window too narrow: variance too large
  Window too wide: possible smoothing of dynamics

Illustration (cont'd): consequence of the use a too wide window $\gamma = 10$ on the ETFE of slide 11



Above plot: the smoothed ETFE at ω_k; Bottom plot: the same with the frequency response of $G_0(e^{j\omega})$ (blue)

Besides trial-and-error coupled with physical insights on $G_0(z)$,

is there another way to select γ?

Yes....

To find this way, note that

$$\begin{aligned}\hat{G}(e^{j\omega}) &= \frac{Y_N(\omega)}{U_N(\omega)} \\ &= \frac{Y_N(\omega)U_N^*(\omega)}{U_N(\omega)U_N^*(\omega)} \\ &= \frac{\sum_{\tau=-\infty}^{+\infty} \hat{R}_{yu}^N(\tau)\, e^{-j\omega\tau}}{\sum_{\tau=-\infty}^{+\infty} \hat{R}_{u}^N(\tau)\, e^{-j\omega\tau}}\end{aligned}$$

where the last step follows from expressions (3.13) and (3.19) in the lecture note.

$$\hat{G}(e^{j\omega}) = \frac{\sum_{\tau=-\infty}^{+\infty} \hat{R}_{yu}^N(\tau)\, e^{-j\omega\tau}}{\sum_{\tau=-\infty}^{+\infty} \hat{R}_{u}^N(\tau)\, e^{-j\omega\tau}} \qquad \textbf{SPA}$$

with

$$\hat{R}_u^N(\tau) = \begin{cases} \frac{1}{N}\sum_{t=0}^{N-1} u(t)u(t-\tau) & for\ |\tau| < N-1 \\ 0 & for\ |\tau| > N-1 \end{cases}$$

$$\hat{R}_{yu}^N(\tau) = \begin{cases} \frac{1}{N}\sum_{t=0}^{N-1} y(t)u(t-\tau) & for\ 0 < \tau < N-1 \\ 0 & elsewhere \end{cases}$$

Interpretation

$\hat{G}(e^{j\omega})$ **can thus be seen as the ratio** $\frac{\hat{\Phi}_{yu}(\omega)}{\hat{\Phi}_u(\omega)}$ **of the approximation** $\hat{\Phi}_{yu}(\omega)$ **of** $\Phi_{yu}(\omega) \triangleq \mathcal{F}(R_{yu}(\tau))$ **and of the approximation** $\hat{\Phi}_u(\omega)$ **of** $\Phi_u(\omega) \triangleq \mathcal{F}(R_u(\tau))$.

This seems logical since

$$\frac{\Phi_{yu}(\omega)}{\Phi_u(\omega)} = \frac{G_0(e^{j\omega})\Phi_u(\omega)}{\Phi_u(\omega)} = G_0(e^{j\omega})$$

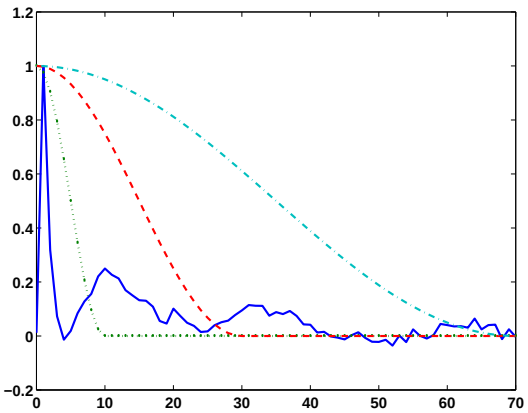
The approximations of the spectra are obtained by taking the Fourier transforms of estimates $\hat{R}_{yu}^N(\tau)$ **and** $\hat{R}_u^N(\tau)$ **of the exact correlation functions.**

Moreover it can be shown that

$$\hat{G}_{sm}(e^{j\omega}) = \frac{\sum_{\tau=-\infty}^{+\infty} w_\gamma(\tau)\hat{R}_{yu}^N(\tau)\, e^{-j\omega\tau}}{\sum_{\tau=-\infty}^{+\infty} w_\gamma(\tau)\hat{R}_{u}^N(\tau)\, e^{-j\omega\tau}}$$

with $w_\gamma(\tau)$ **obtained as the inverse Fourier transform of the frequency window** $W_\gamma(\omega)$

Hamming lag-window $w_\gamma(\tau)$



typical $\hat{R}_{yu}^N(\tau)$ **(solid) together with the Hamming lag-windows** $w_{10}(\tau)$ **(dotted),** $w_{30}(\tau)$ **(dashed) and** $w_{70}(\tau)$ **(dash-dotted).**

$w_\gamma(\tau)$ is a window with width γ: $w_\gamma(\tau) = 0,\ |\tau| > \gamma$

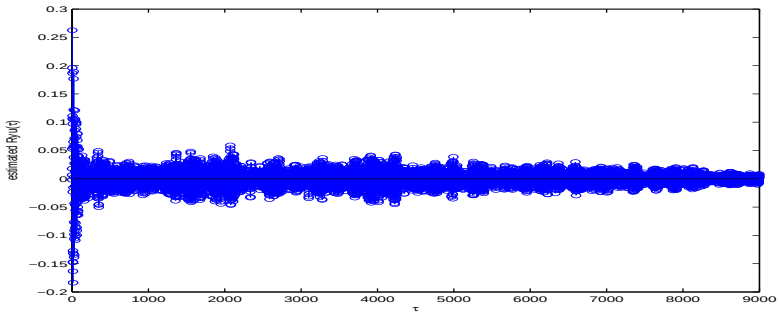
Smoothing corresponds thus to remove from the estimate $\hat{\Phi}_{yu}(\omega)$ of $\Phi_{yu}(\omega)$ the elements of $\hat{R}^N_{yu}(\tau)$ for $\tau > \gamma$

This is relevant since $R_{yu}(\tau) \to 0$ for $\tau \to \infty$ ($G_0(z)$ stable) and since the accuracy $\hat{R}^N_{yu}(\tau)$ is smaller and smaller for increasing values of τ ($\hat{R}^N_{yu}(\tau)$ computed with less data points)

Method for the selection of γ: choose γ such that, for $\tau > \gamma$, $\hat{R}^N_{yu}(\tau)$ are small w.r.t $|\hat{R}^N_{yu}(0)|$ and "less reliable"

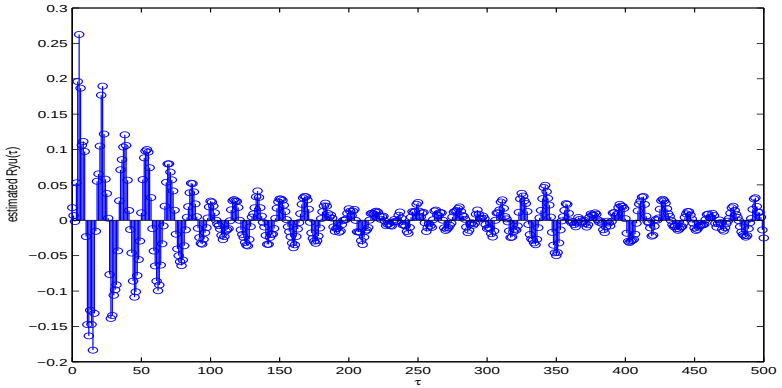
Illustration (cont'd):

We compute $\hat{R}^N_{yu}(\tau)$ with the data generated by the white noise of variance 0.5



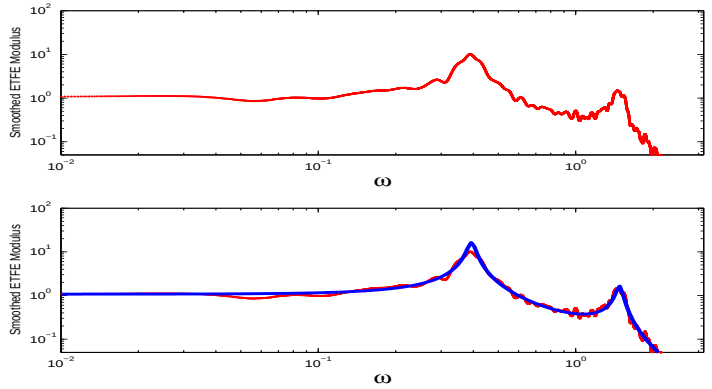
we see the inaccuracy of the estimate: $\hat{R}^N_{yu}(\tau)$ does not tend to 0 for $\tau \to \infty$

Let us focus on the first 500 τ's



we see that, after $\tau = 100$, $\hat{R}^N_{yu}(\tau)$ increases again which is much unlikely for $R_{yu}(\tau) \Longrightarrow$ we select $\gamma = 100$

Obtained smoothed ETFE with $\gamma = 100$



Above plot: the smoothed ETFE at ω_k; Bottom plot: the same with the frequency response of $G_0(e^{j\omega})$ (blue)

Final remarks: drawbacks of ETFE

ETFE gives a discrete estimate of the frequency response of $G_0(e^{j\omega})$ and not the rational transfer function $G_0(z)$

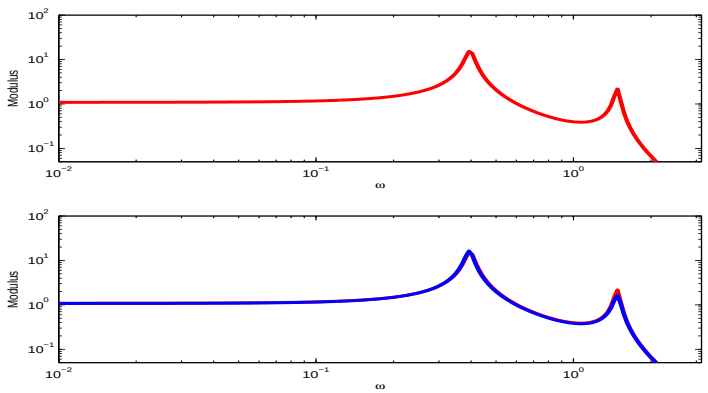
For simulation, for modern control design, such a transfer function is necessary

No information about the noise spectrum $\Phi_v(\omega)$ while this information is important for e.g. disturbance rejection

$\Longrightarrow$ parametric identification (prediction error identification)

- delivers a model of the plant G_0 and information on $\Phi_v(\omega)$
- higher accuracy ($cov(G(e^{j\omega}, \hat{\theta}_N) \approx \frac{n}{N}\frac{\Phi_v(\omega)}{\Phi_u(\omega)}$ with PEI)

Illustration (cont'd): Parametric Identification of $G_0(z)$ with the 1000 first samples of the white noise of variance 0.5



Above plot: frequency response of the identified model; Bottom plot: the same with the frequency response of $G_0(e^{j\omega})$ (blue)

the previous figure has to be compared with the non-smoothed ETFE and the smoothed ETFE

This comparison shows that PEI delivers much better results even with ten times less data points

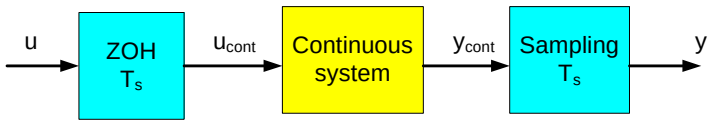
Part V: practical issues when designing the identification experiment

1 Preparatory experiments

- noise measurement on the output
- step response analysis
  - area of linearity
  - time constants
  - static gain
  - delay of the system

Possibilities depend on circumstances

2 Choice of the sampling frequency $\omega_s = \frac{2\pi}{T_s}$



Data for the ETFE with high(est) value of ω_s

Indeed, the higher ω_s, the larger the frequency range that is capured (Shannon theorem)

The ETFE obtained with these data can be represented up to $\frac{\omega_s}{2}$

By inspecting this ETFE, it is then possible to determine the bandwidth ω_b of the system ($\omega_b << \frac{\omega_s}{2}$)

Data for parametric (PEI) identification with smaller ω_s

High ω_s induces numerical problems with parametric identification

Indeed all poles cluster around $z = 1$ since the discrete-time state-space matrix $A_d = e^{A_{cont}\ T_s} \rightarrow I$ when $T_s \rightarrow 0$

Typical choice for parametric identification:

$$10\omega_b < \omega_s < 30\omega_b$$

with ω_b as observed in the ETFE

Data with a smaller ω_s can be obtained

- either by re-collecting data with a smaller ω_s
- or by decimating the data obtained with high ω_s (+anti-aliasing filter)

Remark (actual vs. normalized frequencies):

The model of G_0 identified with data collected with a sampling frequency ω_s contains information up to the Nyquist frequency $\frac{\omega_s}{2}$ (actual frequency)

Considering now the normalized frequency $\omega = \omega_{actual}\ T_s$

We note that the interval $[0\ \frac{\omega_s}{2}]$ (actual frequencies) corresponds to the main interval $[0\ \pi]$ when considering normalized frequencies. Indeed

$$\underbrace{\frac{\omega_s}{2} = \frac{\pi}{T_s}}_{actual\ frequency} \implies \text{normalized } \omega = \pi$$

3 Input signals used for system identification

Finite-power quasi-stationary signals for continuous excitation

- periodic signals (in particular multisines)
- realization of stochastic process ((filtered) white noise or alike)

Trade-off when designing the excitation signal

- the power $\mathcal{P}_u$ / $\Phi_u(\omega)$ should be as high as possible to increase the accuracy of the identified model
- the amplitude of the time-domain signal should be bounded/limited in order not to damage the actuators and in order not to excite the nonlinearities

Multisines

$$u(t) = \sum_{k=1}^{n} A_k\ sin(k\omega_0 t + \phi_k)$$

$\Phi_u(\omega)$ made up of Dirac pulses at the frequencies of the sines in the multisines

the phase shifts ϕ_k can be optimized in order to reduce the maximal amplitude of $u(t)$ without any effect on the power spectrum $\Phi_u(\omega)$

Realization of a stochastic process

$$u(t) = F(z)w(t)$$

with $F(z)$ an user-selected filter and $w(t)$ a white noise of variance σ_w^2

The power spectrum is given by:

$$\Phi_u(\omega) = |F(e^{j\omega})|^2 \sigma_w^2$$

Shaping $\Phi_u(\omega)$ is very easy, but there is no a-priori bound on the amplitude of $u(t)$!!

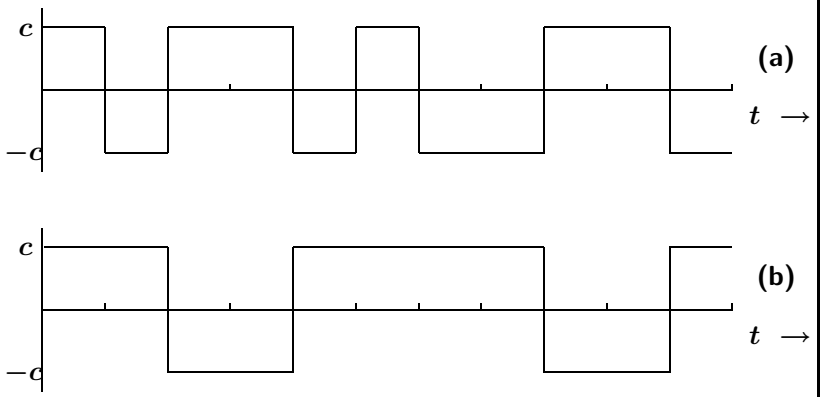
Alternative: Random Binary Sequence (RBS)

$$u(t) = c\ sign\left(w\left(int\left(\frac{t}{\nu}\right)\right)\right)$$

with c the amplitude, $w(t)$ a white noise of variance σ_w^2 and ν the so-called clock period which is an integer such that $1 \leq \nu$

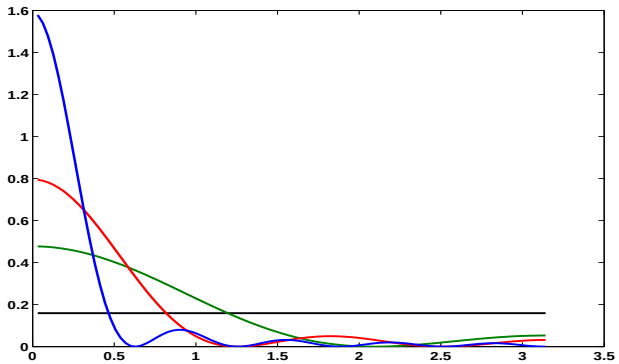
The amplitude of the RBS is either $+c$ or $-c$

The RBS has the maximal power $\mathcal{P}_u = \bar{E}u^2(t) = c^2$ that can be attained by a signal $u(t) \leq c\ \forall t$



(a) Typical RBS with clock period equal to sampling interval ($\nu = 1$);
(b) RBS with increased clock period $\nu = 2$.

Influence of ν on $\Phi_u(\omega)$



Spectrum $\frac{1}{2\pi}\Phi_u(\omega)$ of (P)RBS with basic clock period $\nu = 1$ (black), $\nu = 3$ (green), $\nu = 5$ (red), and $\nu = 10$ (blue).

The power spectrum $\Phi_u(\omega)$ of the RBS is thus shaped via ν:

- $\nu = 1 \Longrightarrow \Phi_u(\omega) = c^2 \ \forall \omega$ i.e. the RBS has the flat power spectrum of a white noise
- For increasing values of ν, the power spectrum $\Phi_u(\omega)$ will be more and more located in low frequencies

less flexibility, but bounded amplitude !!

Another alternative: P(seudo)RBS

- binary signal constructed from a deterministic shift register
- otherwise very similar to RBS

4 Data (pre)processing

- Anti-aliasing filter
- outliers/spike
- Non-zero mean and drift in disturbances; detrending

5 Remarks on unstable systems

Unstable systems can not be identified in open loop

Experiments has to be done with a stabilizing controller C in closed loop:

$$y(t) = \frac{G_0 C}{1 + G_0 C} r(t) + \frac{H_0}{1 + G_0 C} e(t)$$

$$y(t) = \frac{G_0 C}{1 + G_0 C} r(t) + \frac{H_0}{1 + G_0 C} e(t)$$

Since $r(t)$ is independent of $e(t)$, we can excite the closed-loop system via $r(t)$ and identify a model $\hat{T}(z)$ of $\frac{G_0 C}{1+G_0 C}$

A model for the unstable $G_0(z)$ is then

$$\hat{G}(z) = \frac{\hat{T}(z)}{C(z)(1 - \hat{T}(z))}$$