System Identification
Lecture 1

Introduction

Paul Van den Hof

Control Systems Group
Eindhoven University of Technology

dutch institute of systems and control

Technische Universiteit Eindhoven University of Technology

Contents

Schedule

Sep 14	PVdH	Introduction; concepts
Sep 21	PVdH	Prediction error identification
Sep 28	PVdH	Asymptotic analysis
Oct 5	GB	Bayesian estimation; machine learning
Oct 12	JS	Frequency domain identification
Oct 19	JS	Nonlinear identification
Oct 26	PVdH	Closed-loop identification
Nov 2	PVdH	Dynamic networks

- Notion of model common in many branches of science
- Models come in different shapes and formats
- They intend to replace/reflect the behaviour of (physical) systems

Model-Based Engineering is the single dominant paradigm that underlies the design, operation and maintenance of engineering systems.

interacting robots; nano-positioning systems; smart (energy) grids; autonomous vehicles; (petro)chemical production plants;.....

In systems where dynamics is present and important, the systems and control field has a role to play with its core expertise:

- Modelling tools that are (generically) valid across disciplines
- Theory and tools for analyzing interconnection of systems (e.g. feedback)
- A framework for dealing with uncertainty and creating robustness

• R.M. Murray (Ed.) (2003). Control in an information rich world: Report of the panel on future directions in control, dynamics and systems. SIAM. Available at `http://cds.caltech.edu/~murray/cdspanel`

Within the systems and control domain:
models of dynamical systems are used for the purpose of "Model-Based X":

- control design, optimization
  feedback, feedforward, distributed
- prediction, simulation
- diagnosis, fault detection
- monitoring, soft sensing
- system (re)design

In a huge variety of technology fields and societal challenges.

• F. Lamnabhi-Lagarrigue, A. Annaswamy, S. Engell, A. Isaksson, P. Khargonekar, R. Murray, H. Nijmeijer, T. Samad, D. Tilbury and P. Van den Hof (2017). Systems & Control for the Future of Humanity. Research agenda: current and future roles, impact and grand challenges. *Annual Reviews in Control, 43*, pp. 1-64, 2017.

Modelling of physical systems:

(a) First principles (physical) modelling

- Conservation laws, mass and energy balances, Newton's law,...

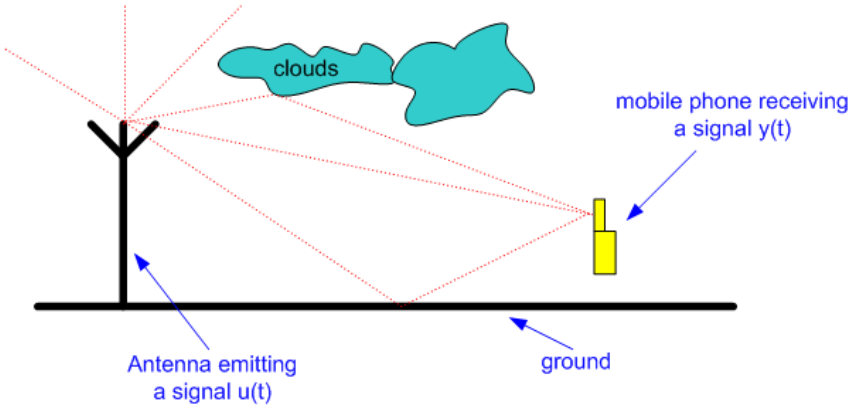
(b) System identification

- modelling from data

Why system identification?

- physical modelling too hard/unreliable/complex/uncertain
- modelling of emergent behaviour
- identification for model verification, parameter estimation

Example 1: Channel identification in mobile telephony



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

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

$$\Longrightarrow \text{distorted signal}$$

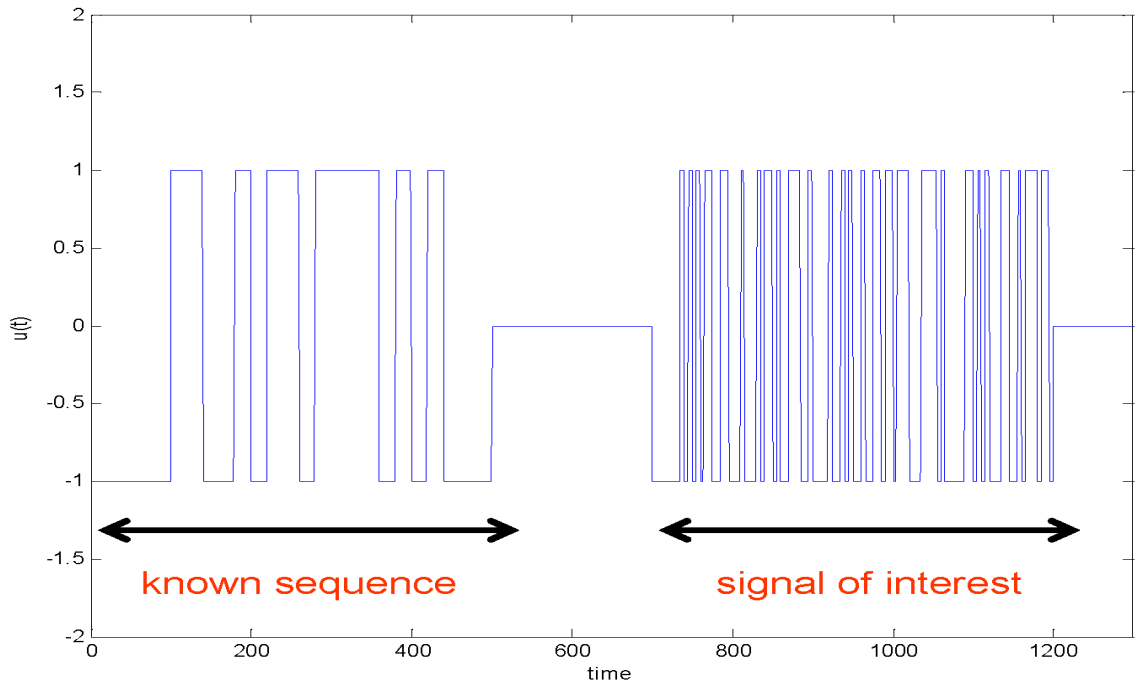
A model of the so-called channel is required to reconstruct $u(t)$ from the distorted $y(t)$

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

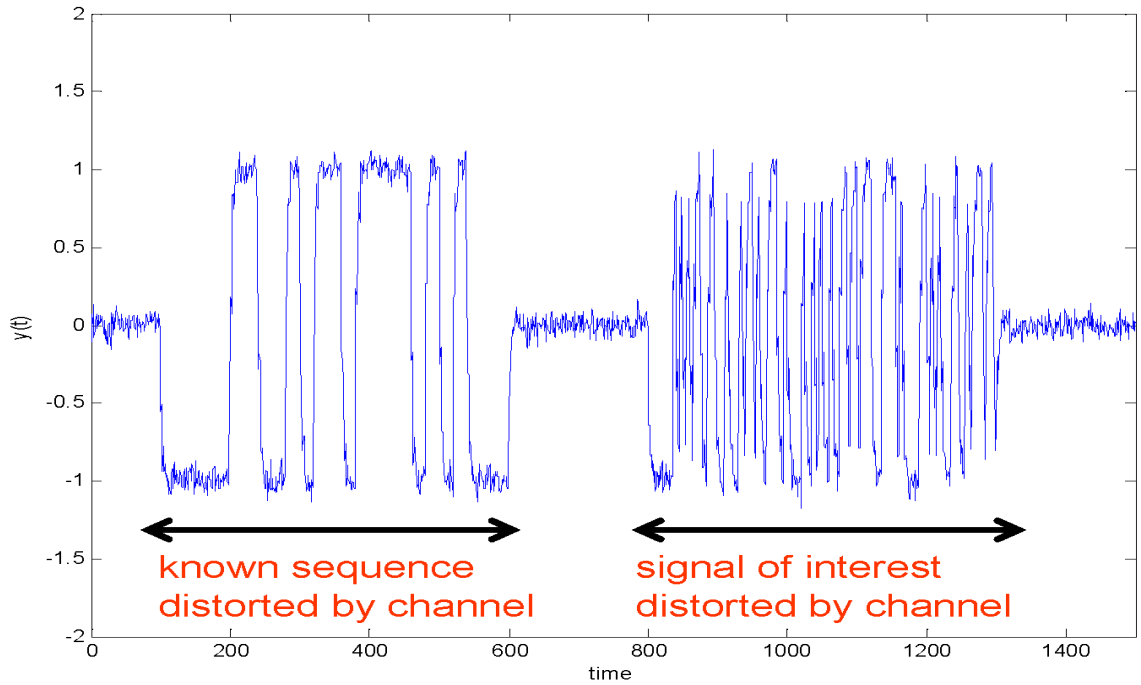
The model is identified at each received call

How to proceed?

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



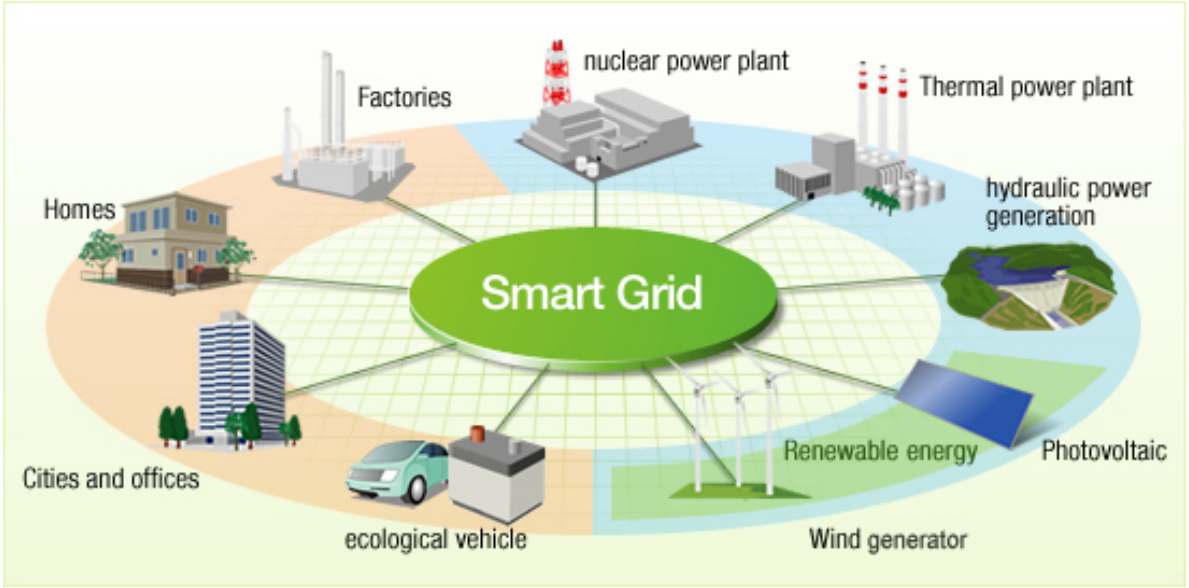
Both the known sequence and the signal of interest are distorted by the channel



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 $y_{interest}(t)$

Example 2: Smart grids



(courtesy of hitachi.com)

- Real-time balancing of supply and demand of power
- Dynamics of power system changes dependent on mix of generation (inertia of traditional power plants versus fast-changing renewables)

Future requirements on engineering systems:

Handling of highly complex / interacting / distributed systems that

operate autonomously / with changing objectives / in a "learning" mode / adapting to changing circumstances

and maintain a verifiable high performance.

Some required future capabilities of models:

1. Accuracy assessment
   on-line assessment of model validity
2. Adaptability
   flexible on-line updating of models (dynamics and interconnection structure)
3. Active data-driven learning
   demands on accuracy, autonomy, robustness $\rightarrow$ active probing for information

All these concepts relate to data-driven modelling which becomes an integral part in virtually all complex engineering systems

What is a system / model?

In most fields of science determined by a set of equations (conservation balances) that govern the underlying physical phenomena, and restrict the time-behaviour of related variables.

- ODE (ordinary differential equations)
- PDE (partial differential equations)
- ADE (algebraic differential equations)

with linear and/or nonlinear dynamics, and
in either continuous dynamics or switching dynamics or a combination (hybrid systems).

State-space representation

$$\begin{aligned} \dot{x}(t) &= f(x(t), u(t)), \quad x(0) = x_0 \\ y(t) &= g(x(t), u(t)) \end{aligned}$$

Linear case:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \quad x(0) = x_0 \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

Discrete-time:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \quad x(0) = x_0 \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

Steady state behaviour characterized by transfer function $\mathbb{C} \to \mathbb{C}$

$$H(z) = C(zI - A)^{-1}B + D$$

In this course:

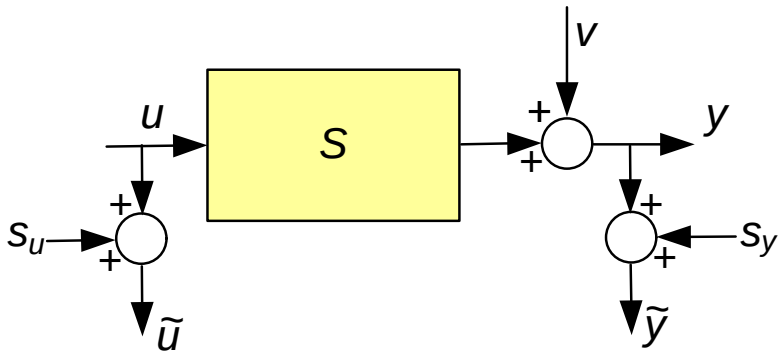
1. Linear discrete-time continuous dynamics
   (PE methods - PVdH)
2. Linear continuous-time continuous dynamics
   (f-domain ID - JS)
3. Nonlinear, continuous dynamics
   (non-linear ID - JS)

Identification setup

In systems and control theory, there is a general need to think in terms of cause/effect relationships:

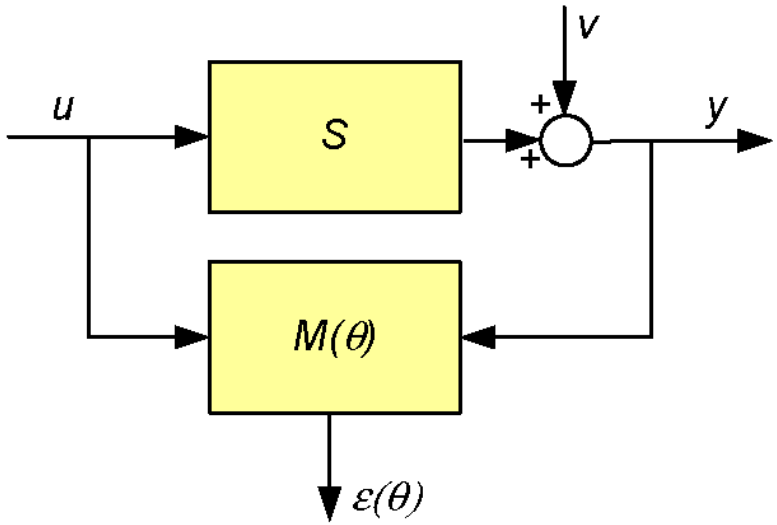
Signals to be distinguished:

- measurable input signals
- (measurable) output signals
- non-measurable disturbances
- sensor noise



Choice of input/output may not be unique

Prediction error methodology in short



Model class:

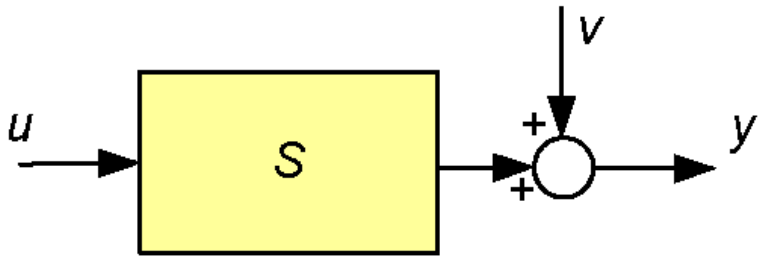
$$y(t) + a_1 y(t-1) + a_2 y(t-2) =$$
$$= b_1 u(t-1) + b_2 u(t-2) + \varepsilon(t)$$

Parameter: $\theta = (a_1 \;\; a_2 \;\; b_1 \;\; b_2)^T$

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

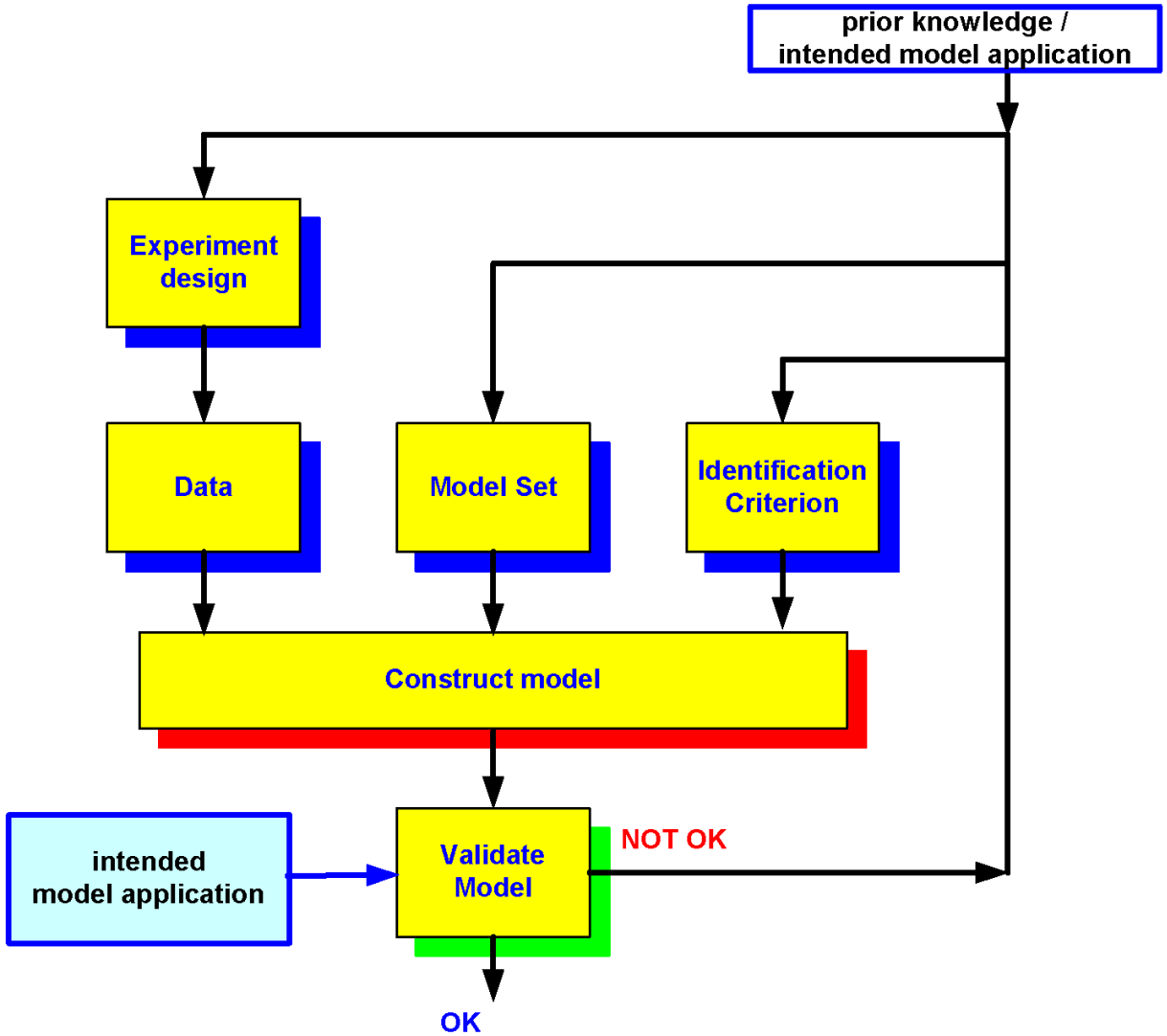
Two forms of appearance of models

(a) Model as reflection of the data generating system



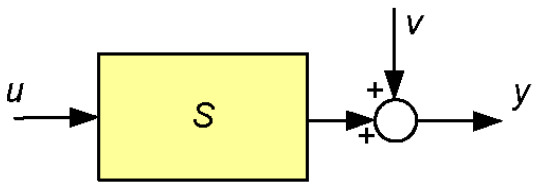
(b) Model used for identification

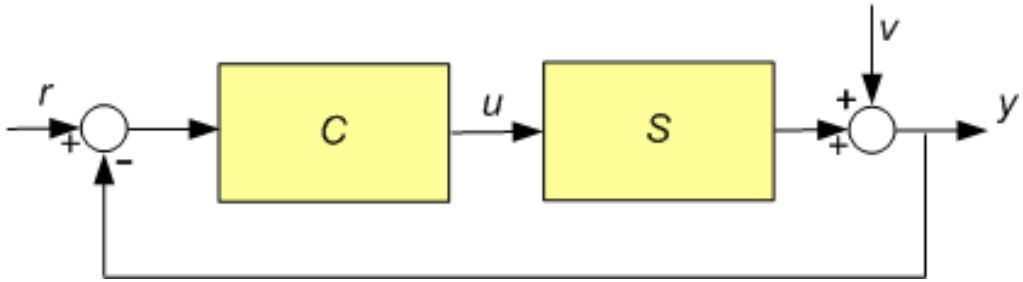
Classical system identification procedure (Ljung, 1987)



Experiment Design

- Inputs to be used for deliberate excitation of dynamics
- Which frequency content of the input signals?
- Open-loop (free excitation)
  or closed-loop (controlled excitation)?





- Periodic signals (sum of sinusoids) ⇓ f-domain identification

or

non-periodic signals ⇓ t-domain identification

Model set

- Frequency response (graphical, non-parametric) ⇐
- Pulse response / Step response
- Linear difference / differential equation ⇐
  (transfer function / state-space representation)
- Nonlinear difference / differential equation ⇐
- First principles, black box or combination .....

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

In this course mainly:
SISO black box, with some indicated extensions.

Identification Criterion

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

- Least squares (LS)
- Instrumental variables (IV)
- Maximum likelihood (ML)
- Bayesian methods
- ..... algorithm-driven (subspace identification methods)

dependent on prior knowledge on stochastic properties of the signals

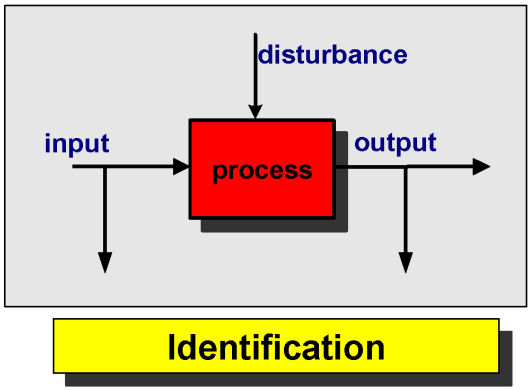
History

- 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 a mature field in Ljung (1987; 1999)
  and 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, ...).
- Extensions towards nonlinear models and machine learning (data analytics)

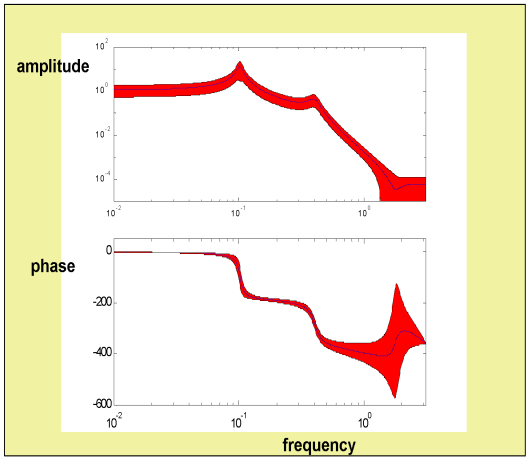
Current topics of development

- From SISO to MIMO (large scale)
- Black box and first principle models
- From linear to nonlinear models
- From fixed model sets to automated model sets (machine learning)
- From MIMO to structured dynamic networks
- From Model-based X to data-based X (big data);

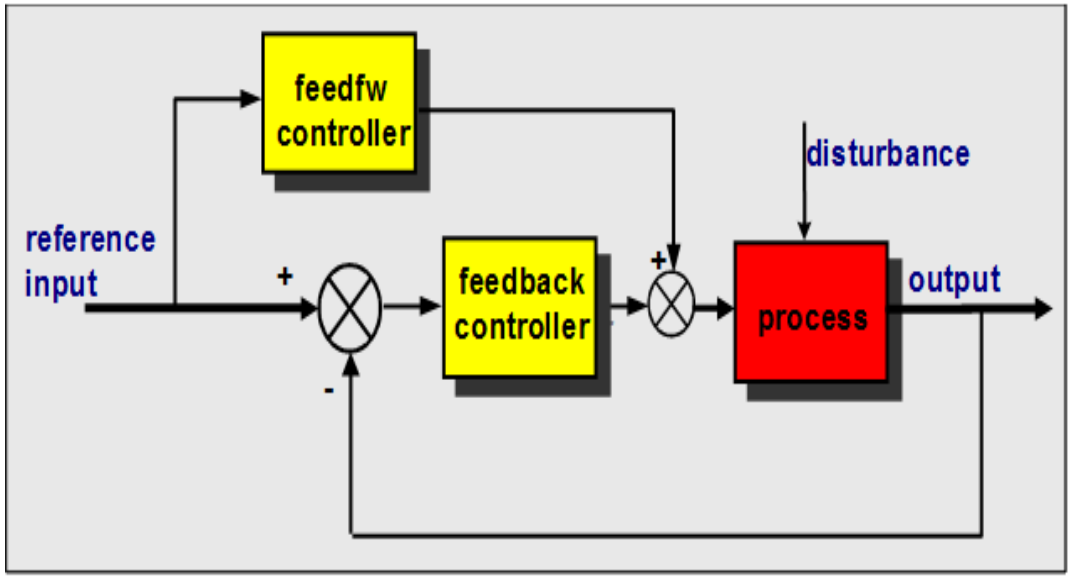
System identification for (robust) control



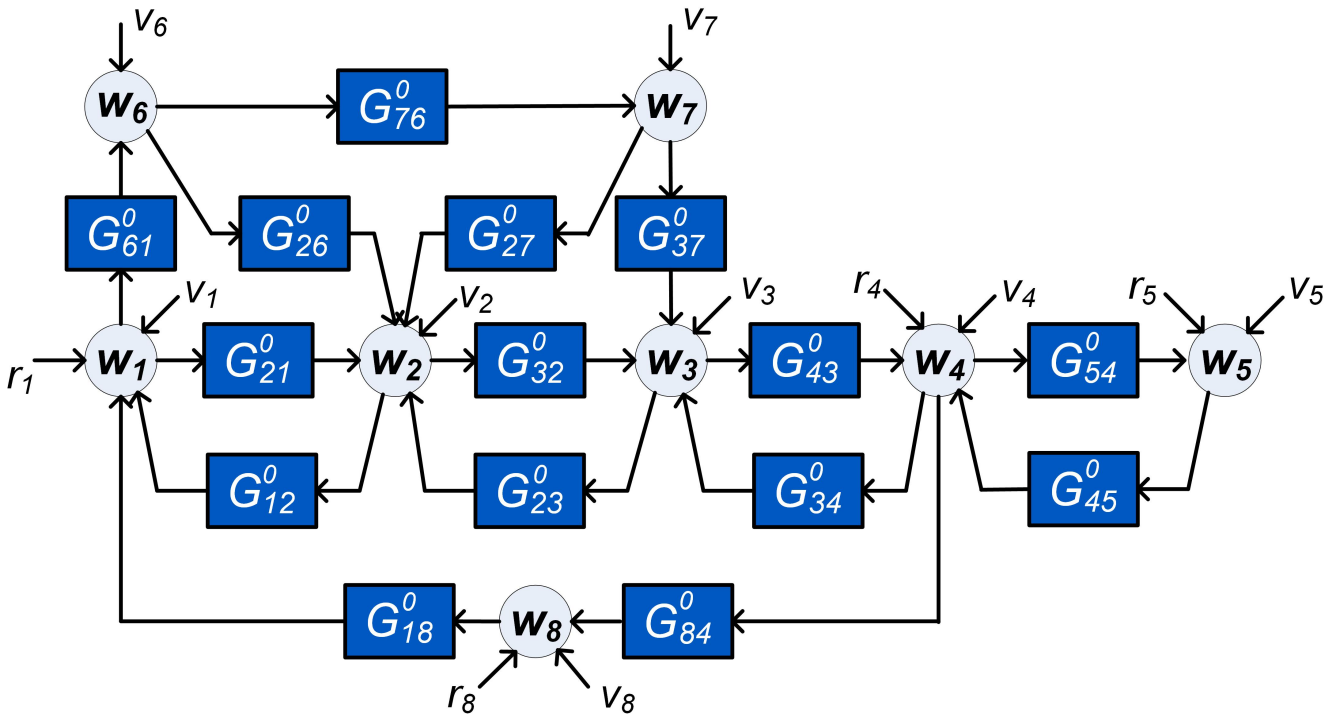
Data → Model



Model → Controller



System identification in dynamic networks



- Identification of one or more (local) modules
- What to measure/model for consistency/minimum variance?
- Identify structure / topology
- Identifiability / uniqueness?

Estimation

Estimator $\hat{\theta}_N$ of θ_0, is a mapping from $\{u(t), y(t)\}_{t=1,\dots N}$ to $\hat{\theta}_N$.
If data contains random variables, then $\hat{\theta}_N$ is a random variable.

Concepts:

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

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

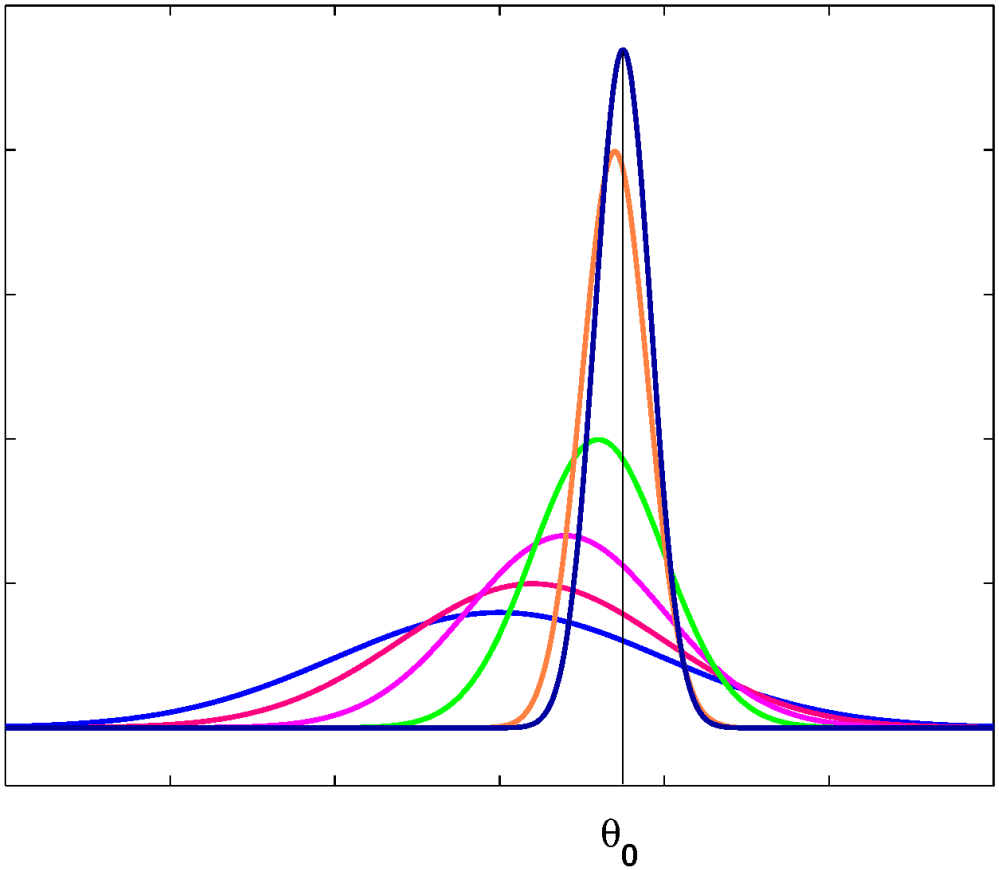
- $Pr[\lim_{N\to\infty} \hat{\theta}_N = \theta_0] = 1$
- $\hat{\theta}_N \to \theta_0$ with probability 1 voor $N \to \infty$.

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

d. *Asymptotic normal:* $\hat{\theta}_N \in As\mathcal{N}(\theta^*, P_\theta)$

Consistency = Asymptotic unbiased and variance tends to 0

Probability density function (pdf) of a consistent estimate



$f(\hat{\theta}_N)$ for increasing values of N

Recap on discrete-time signals and systems

- Discrete-time Fourier transform
- Power spectrum
- Stochastic processes
- Quasi-stationary signals
- Discrete-time systems

For deterministic signals

$$u_d(k) = u(kT_s)$$

Signal Power:

$$\mathcal{P}_u := \frac{1}{N} \sum_{k=0}^{N-1} u_d(k)^2$$

Discrete-time Fourier transform (DTFT) (infinite time):

$$
\begin{aligned}
U_s(\omega) &= \sum_{k=-\infty}^{\infty} u_d(k) e^{-i\omega k T_s} \\
u_d(k) &= \frac{T_s}{2\pi} \int_{2\pi/T_s} U_s(\omega) e^{i\omega k T_s} d\omega,
\end{aligned}
$$

with $\omega_s := \frac{2\pi}{T_s}$ the sample frequency (rad/sec). For a finite $|U_s(\omega)|$ this requires that $\sum |u_d(k)| < \infty$.

$U_s(\omega)$ is:

- continuous function in ω
- periodic with period ω_s

Discrete Fourier Transform (DFT) (finite-time):

$$\begin{aligned} U_N(\omega) &= \sum_{k=0}^{N-1} u_d(k) e^{-i\omega k T_s}. \\ u_d(k) &= \frac{1}{N} \sum_{\ell=0}^{N-1} U_N(\frac{\ell \omega_s}{N}) e^{i\frac{2\pi\ell}{N}k}. \end{aligned}$$

DFT: $\{U_N(\omega),\ \omega = \frac{\ell}{N}\omega_s,\ \ell = 1, \cdots N\}$, N points in ω-domain.
Because of symmetry: $U_N(-\omega) = U_N(\omega)^*$.

Note: $U_N(\omega)$ takes values over a continuous range of ω, but only its values in the Fourier grid are required for reconstructing $u_d(k)$.

Signal power

General expression for power: $\mathcal{P}_u = \frac{1}{\omega_s} \int_{\omega_s} \Phi_u(\omega) d\omega$

with power spectrum $\Phi_u(\omega)$.

For general finite-time signals: $\Phi_u(\omega) = \frac{1}{N} |U_N(\omega)|^2$

referred to as the periodogram of the signal.

Equivalent power expression for finite time signals:

$$\mathcal{P}_u = \frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{N} |U_N(\frac{2\pi k}{N})|^2$$

Stationary stochastic processes

Power of a (zero-mean) process x: $\mathbb{E}x^2(k)$

- (Auto)-correlation function

$$R_x(\tau) = \mathbb{E}[x(k)][x(k-\tau)]$$

- Power spectrum

$$\Phi_x(\omega) = \sum_{k=-\infty}^{\infty} R_x(k) e^{-i\omega k T_s}$$

$$\Rightarrow \quad R_x(0) = \mathbb{E}x^2(k) = \frac{1}{\omega_s} \int_{\omega_s} \Phi_x(\omega) d\omega$$

Power spectrum is Fourier transform of auto-correlation function

Quasi-stationarity

$$y(k) = w(k) + v(k)$$

w deterministic signal
v stationary stochastic process

With repetition of experiments: w unchanged, v not.
Consequence: y *non-stationary* stochastic process!

Introduction of $\bar{\mathbb{E}}$:

$$\bar{\mathbb{E}} y(k) := \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathbb{E} y(k)$$

- time-averaging operates on deterministic part
- $\mathbb{E}$ operates on stochastic part

Resulting notions

- *(Auto-)correlation function*:

$$R_y(\tau) := \bar{\mathbb{E}}[y(k)y(k-\tau)]$$

- *Cross-correlation function*:

$$R_{yu}(\tau) := \bar{\mathbb{E}}[y(k)u(k-\tau)]$$

- *Power spectrum*:

$$\Phi_y(\omega) := \sum_{\tau=-\infty}^{\infty} R_y(\tau)e^{-i\omega\tau}$$

- *Cross power spectrum*:

$$\Phi_{yu}(\omega) := \sum_{\tau=-\infty}^{\infty} R_{yu}(\tau)e^{-i\omega\tau}$$

Signals for which these expressions exist: *quasi-stationary*.

Consequence:

$$\bar{\mathbb{E}} y^2(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_y(\omega) d\omega$$

Power spectrum of quasi-stationary signals

Let

$$y(k) = w(k) + v(k)$$

with w quasi-stationary deterministic, and v stationary stochastic with $\mathbb{E} v(k) = 0 \ \forall k$, then

$$\Phi_y(\omega) = \Phi_w(\omega) + \Phi_v(\omega)$$

Proof follows from the observation that $\bar{\mathbb{E}}[w(k)v(k-\tau)] = 0$

Discrete-time systems

$$y(t) = \sum_{k=0}^{\infty} g(k)u(t-k) \qquad t = 0, 1, 2, \cdots$$

$$\{g(k)\}_{k=0,1,\cdots} \quad \text{puls-response}$$

- Linearity $\Rightarrow$ convolution
- Time-invariance $\Rightarrow$ $g(k)$ time-independent
- Causality $\Rightarrow$ $g(k) = 0,\ k < 0$
- discrete-time

Output signal is determined through

- puls-response
- initial conditions; often chosen $u(t) = 0,\ t < 0$

Shift operators

$$\begin{aligned} q\ u(t) &= u(t+1) \\ q^{-1}\ u(t) &= u(t-1) \end{aligned}$$

$$y(t) = \sum_{k=0}^{\infty} g(k) q^{-k}\ u(t) = G(q)\ u(t)$$

Transfer(-operator): $G(q) = \sum_{k=0}^{\infty} g(k)\ q^{-k}$

Transfer function (formal) $\mathbb{C} \to \mathbb{C}$: $G(z) = \sum_{k=0}^{\infty} g(k)\ z^{-k}$.

System is finite dimensional $\leftrightarrow$ G is rational function in z

Stability

G BIBO stable if

$$\sum_{k=0}^{\infty} |g(k)| < \infty$$

- $|u(t)| \leq c_1 \quad \Rightarrow \quad |y(t)| \leq c_2$
- $G(z)$ analytic (no poles) on/outside unit circle $z = e^{i\omega}$.

Frequency response $G(e^{i\omega})$

If $u(t) = cos\ \omega t$, then

$$y(t) = |G(e^{i\omega})| \cdot cos(\omega t + \phi)$$

$$\phi = arg[G(e^{i\omega})]$$

$G(e^{i\omega}),\ 0 \leq \omega \leq \pi$ delivers discrete Bode plot.

Processing of signals:

If $y(t) = G(q)u(t)$,
$u(t)$ quasi-stationary, then

$$\begin{aligned} \Phi_{yu}(\omega) &= G(e^{i\omega})\Phi_u(\omega) \\ \Phi_y(\omega) &= |G(e^{i\omega})|^2 \cdot \Phi_u(\omega) \end{aligned}$$

Summary

- Importance of models in model-based engineering
- Model derivation either through first principles or data-driven
- General identification procedure
- Estimation concepts: bias and variance
- Stationary stochastic process to model random signals
- Quasi-stationary signals to combine det+stoch signals
- Consult material of Chapter 2 if necessary

Prediction models

$$y(t) = G(q)u(t) + v(t)$$

Additive disturbance on the output.

Disturbance signal v:

> Measurement noise; non-measurable input signals; effects of linearization; $\cdots$

Noise model:
v is a stationary stochastic process, with $\mathbb{E}v(t) = 0 \quad \forall t$, and a spectral density $\Phi_v(\omega)$ that is smooth (finite dimensional).

Spectral factorization theorem

If the stochastic process v is generated through a finite dimensional system, i.e. $v(t) = N(q)e(t)$ with N a rational and stable transfer function and e a stationary white noise, then there exists a unique decomposition

$$\Phi_v(\omega) = H(e^{i\omega})\sigma_e^2 H(e^{i\omega})^* = \sigma_e^2 \cdot |H(e^{i\omega})|^2$$

with:

- $H(q)$ and $H^{-1}(q)$ are stable,
- H is monic, i.e. $\lim_{z\to\infty} H(z) = 1$,
- $\mathbb{E}e^2(t) = \sigma_e^2$.

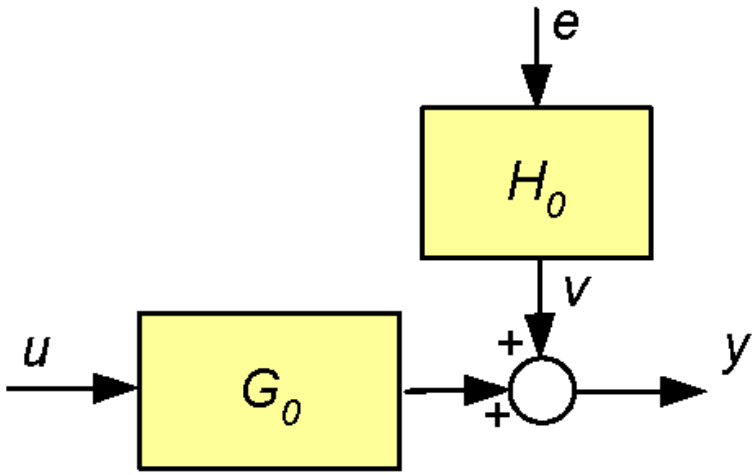
We can then write: $v(t) = H(q)e(t)$, with (H, σ_e^2) as a unique representation of the disturbance process v.

A system/model is therefore characterized by the relation

$$y(t) = G(q)u(t) + H(q)e(t)$$

with H monic, stable and stably invertible (minimum-phase).

All poles and zeros of H are within the unit circle.



Prediction

Given a dynamical system

$$y(t) = G(q)u(t) + H(q)e(t)$$

and given observations

$$\{y(s), s \leq t-1; u(s), s \leq t\} \quad i.e. \quad y^{t-1}, u^t$$

how can the future value $\boldsymbol{y(t)}$ be predicted?

Write $y(t)$ as much as possible in terms of past values of y, u:

$$\begin{aligned} y(t) &= G(q)u(t) + H(q)e(t) \\ &= G(q)u(t) + [H(q) - 1]e(t) + e(t) \end{aligned}$$

Substitute in the second term:

$$e(t) = H(q)^{-1}[y(t) - G(q)u(t)]$$

Then:

$$\begin{aligned} y(t) &= G(q)u(t) + [H(q)-1]H^{-1}(q)[y(t)-G(q)u(t)] + e(t) \\ &= H^{-1}(q)G(q)u(t) + [1 - H^{-1}(q)]y(t) + e(t). \end{aligned}$$

Analyzing the properties of this equation is supported by structural properties of $H^{-1}(q)$.

Property of H^{-1}:

For

$$H(z) = 1 + \sum_{k=1}^{\infty} h(k) z^{-k}$$

there holds (since H is stable and inversely stable)

$$\begin{aligned} \frac{1}{H(z)} &= \frac{1}{1 + h(1)z^{-1} + h(2)z^{-2} + \cdots} \\ &= 1 - h(1)z^{-1} + \cdots \end{aligned}$$

$(1 + \textit{strictly proper})^{-1} = (1 + \textit{strictly proper})$

As a consequence:

- $H^{-1}(q)$ is proper (causal), and
- $1 - H^{-1}(q)$ is strictly proper.

As a result:

$$y(t) = \underbrace{H^{-1}(q)G(q)u(t)}_{\text{known at } t} + \underbrace{[1 - H^{-1}(q)]y(t)}_{\text{known at } t-1} + \underbrace{e(t)}_{\textit{unknown}} .$$

The random variable $y(t)$ conditioned on y^{t-1}, u^t has a pdf that is induced by f_e, with $\mathbb{E}_{f_e} = 0$.
One option for the choice of the "best" predictor is to choose

$$\hat{y}(t|t-1) := \mathbb{E}\{y(t) \mid y^{t-1}, u^t\}$$

which can be calculated on the basis of $\mathbb{E}[e(t)|y^{t-1}, u^t] = 0$.

One-step-ahead prediction

For a dynamical model $y(t) = G(q)u(t) + H(q)e(t)$ the one-step-ahead prediction $\hat{y}(t|t-1) := \mathbb{E}\{y(t)|y^{t-1}, u^t\}$ is given by

$$\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + [1 - H^{-1}(q)]y(t)$$

At the same time:

$$y(t) = \hat{y}(t|t-1) + e(t)$$

The one-step ahead predictor of a dynamic model predicts the output of the model up to the white noise term $e(t)$, called the innovation process.

Prediction in a state space setting

General state space model:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + w(k), \quad x(0) = x_0,\ cov(x_0) = P \\ y(t) &= Cx(t) + Du(t) + v(k) \end{aligned}$$

with w, v white noise processes, can be converted to the innovation form

$$\begin{aligned} \xi(t+1) &= A\xi(t) + Bu(t) + K(t)e(t), \quad \xi(0) = \xi_0 \\ y(t) &= C\xi(t) + Du(t) + e(t) \end{aligned}$$

with

$$\bar{\mathbb{E}}\{y(t) \mid \xi^t, u(t)\} = C\xi(t) + Du(t),$$

and e the innovation process.

The optimal -time-varying- Kalman gain $K(t)$ is obtained through Kalman filter theory.

Solution for $t \to \infty$ is equivalent to transfer function expression.

Example 1 - Predictor for Moving Average (MA) model

$$y(t) = e(t) + ce(t-1)$$

$$H(z) = 1 + cz^{-1}; \qquad H^{-1}(z) = \frac{1}{1 + cz^{-1}}$$

$$1 - H^{-1}(z) = \frac{cz^{-1}}{1 + cz^{-1}}$$

So

$$\hat{y}(t|t-1) = \frac{cq^{-1}}{1 + cq^{-1}} y(t)$$

or

$$\begin{aligned} (1 + cq^{-1})\hat{y}(t|t-1) &= cq^{-1}y(t) \\ \hat{y}(t|t-1) + c\hat{y}(t-1|t-2) &= cy(t-1). \end{aligned}$$

Note: $\mathbb{E}y(t) = 0$ but still the prediction of $y(t)$ is unequal 0! (y^{t-1} carries information on $y(t)$).

Example 2 - Predictor for Output error (OE) model

$$y(t) = \frac{b_1 q^{-1}}{1 + a_1 q^{-1}} u(t) + e(t)$$

$$G(q) = \frac{b_1 q^{-1}}{1 + a_1 q^{-1}}; \qquad H(q) = 1.$$

So:

$$\begin{aligned} \hat{y}(t|t-1) &= H^{-1}(q)G(q)u(t) + (1 - H^{-1}(q))y(t) \\ &= G(q)u(t). \end{aligned}$$

Predictor is only driven by input signals.

Summary

- Framework for linear time-invariant finite-dimensional models with disturbances
- Models are represented by G, H and σ_e^2
- Models induce prediction properties