Identification

University of Ljubljana, Faculty of Electrical Engineering

2nd Cycle Postgraduate Study Programme, Year 1

Study programme module A

Summer semester 2024/25

Sašo Blažič

UNIVERZA
V LJUBLJANI

FE

Fakulteta
za elektrotehniko

Laboratorij za Avtomatiko
in Kibernetiko

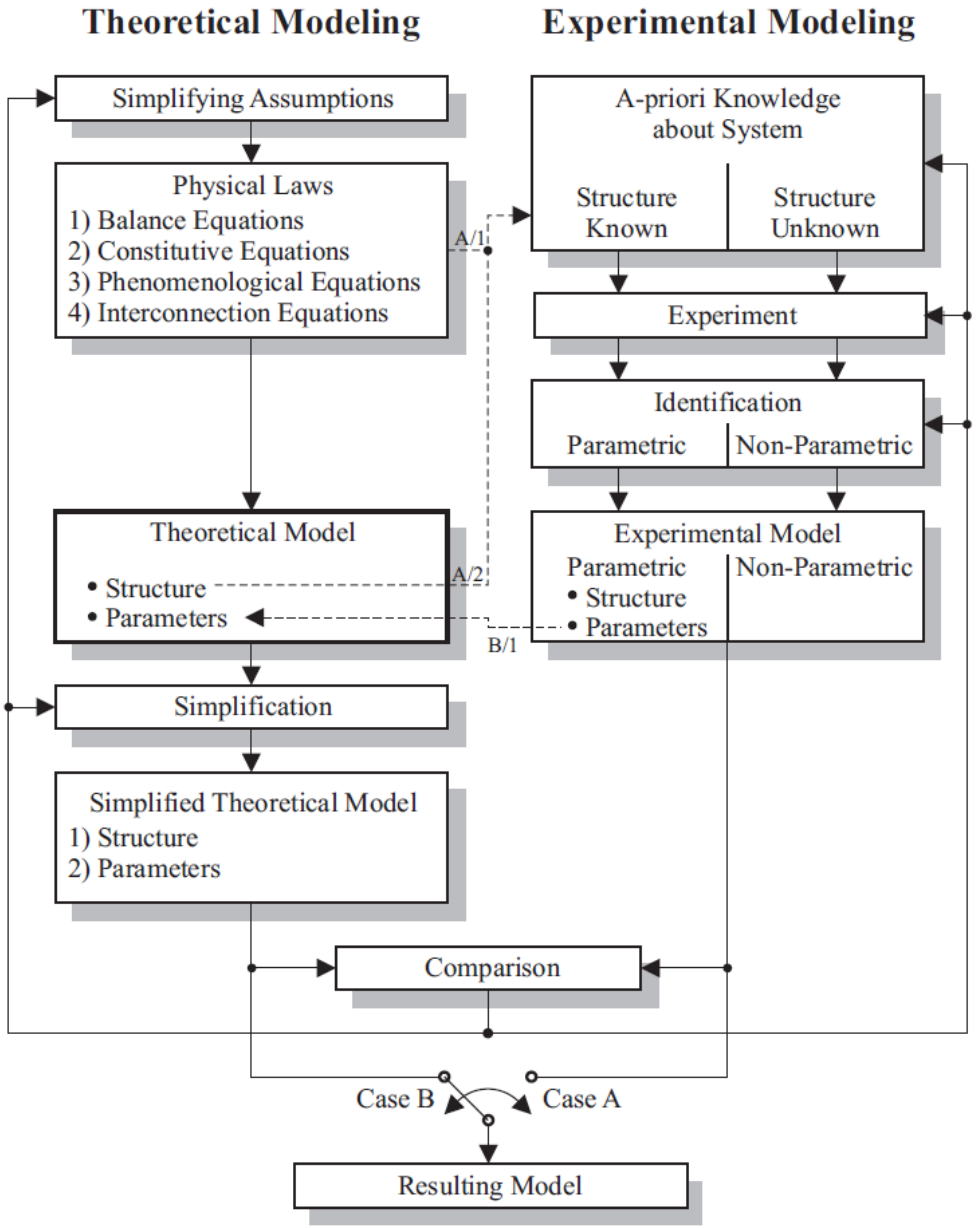
Models as an abstraction of reality

- Model building based on system observation and study of its properties is an essential ingredient of modern sciences
- Models can have more or less formal character but the main idea is always to try to find a pattern in the observations or extract features to build the model (with the same properties as a system)
- A system – a confined arrangement of mutually affected entities (processes) while a process is defined as the conversion and/or the transport of material, energy, and/or information (DIN 66201)
- We will only deal with *mathematical models* that can describe the behaviour of an arbitrary system (technical, biological, economic etc.) – this is the fundament of modern *system theory*
- A mathematical model – obtained by mathematical modelling:
  - Theoretical – first principles
  - Experimental – identification

Identification of dynamical systems

- Alternative way of modelling – based on analysis of measured signals – therefore the name **experimental modelling** or **system identification** or simply **identification**
- The goal of identification – to determine system model based on available measurement data
- Very attractive approach – the model can be (in theory) obtained without the knowledge of any system properties (black box)
- However, the measured data are (almost) always inaccurate and imprecise due to problems with sensors and measurement systems
- Another problem – models are always incomplete – real systems are usually much more complex than the simplified models
- Searching for an ideal model is futile – what is needed is a model that is **useful** – it fulfils its **purpose**

Theoretical modelling vs. identification



Theoretical modelling (ThM) vs. identification

- The result of ThM – system of ODEs (ordinary differential equations) and/or PDEs (partial differential equations) – if able to solve it, we obtain a theoretical model (structure and parameters)
- Often too comprehensive and/or complex – needs simplification before being useful
- Even if explicit solution not possible $\rightarrow$ the equations give important information about system structure:
  - balance equations always linear
  - phenomenological equations linear on a broad operational region
  - constitutive equations (or physical or chemical equations of state) often introduce nonlinearities
- Identification – a mathematical model of a process is based on measurements – any a priori knowledge obtained with ThM or data analysis is crucial for determining model structure (reduces to system order with LTI systems)
- Next steps:
  - performing the measurements and signal processing
  - identification algorithm – processes measurements – the result is usually an input-output model
- Input signals can be natural (taken during normal behaviour) or artificial test signals
- Identification of parametric or non-parametric models can be done (based on the model purpose)

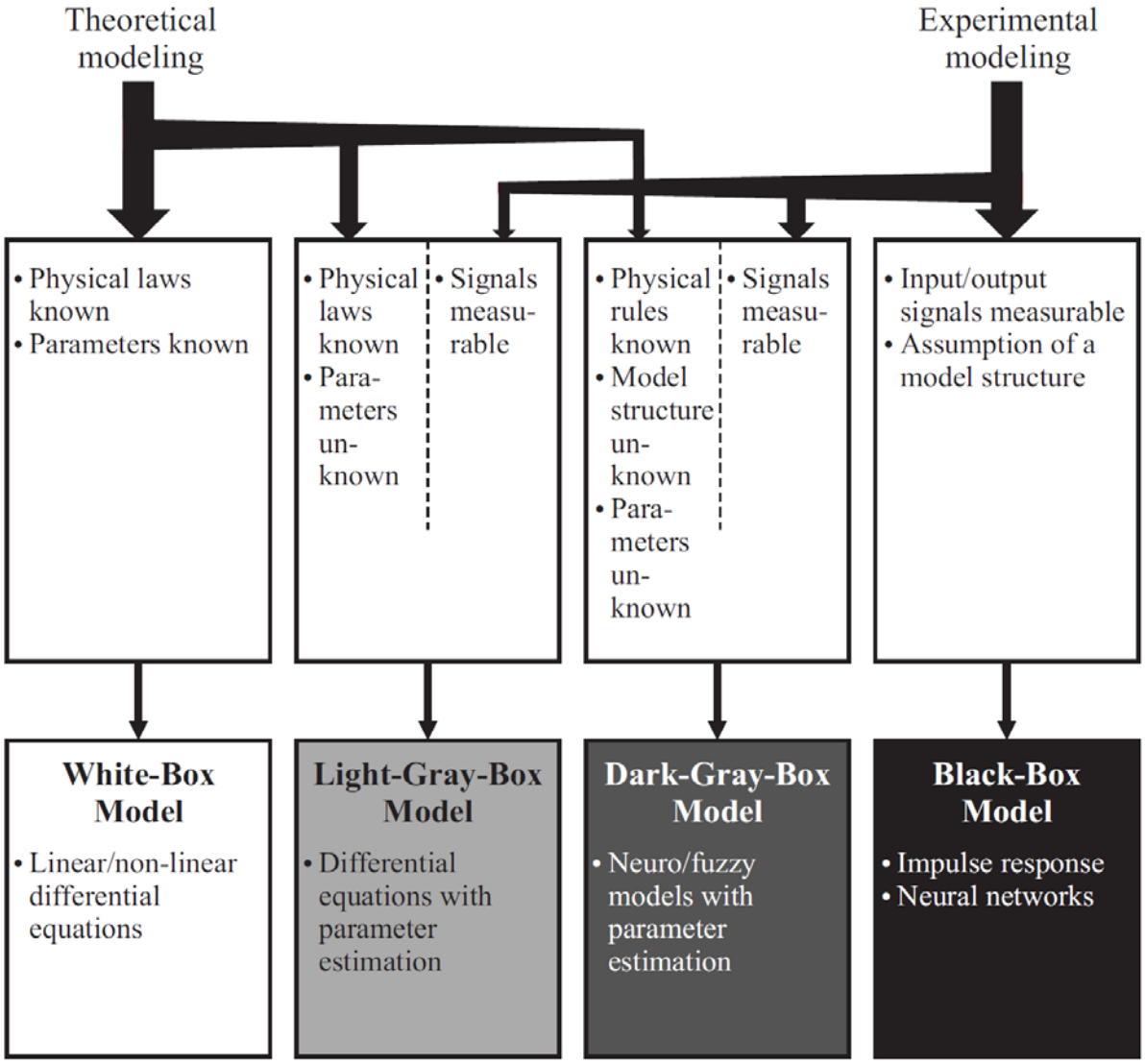
Theoretical modelling vs. identification

- A theoretical model includes functional relation between physical and model parameters
- Experimental model – parameters are usually numerical values without an obvious relation to the physical parameters
- But – experimental models usually describe system behaviour more accurately or at least with lower cost
- Theoretical and experimental modelling – complementary:
  - A feedback loop in the flowchart based on model comparison
  - Modelling essentially an iterative procedure
  - Combination of both approaches is ideal:
    - Theoretical modelling → model structure
    - Identification → model parameters

Theoretical modelling vs. identification

Theoretical modelling	Identification
Model structure follows from laws of nature	Model structure must be assumed
Modelling of the input/output behaviour as well as the internal behaviour	Only the input/output behaviour is identified
Model parameters are given as functions of system properties	Model parameters are “numbers” only, in general no functional dependency to system properties known
Model is valid for the entire class of processes of a certain type and for different operating conditions	Model is only valid for the investigated system and within operating limits
Model coefficients are not known exactly	Model coefficients are more precise for the given system within the operating limits
Models can be built for non-existing systems	Model can only be identified for an existing system
The internal behaviour of the system must be known and must be describable mathematically	Identification methods are independent of the investigated system and can thus be applied to many different systems
Typically a lengthy process which takes up much time	Fast process if identification methods exist already

Theoretical modelling vs. identification

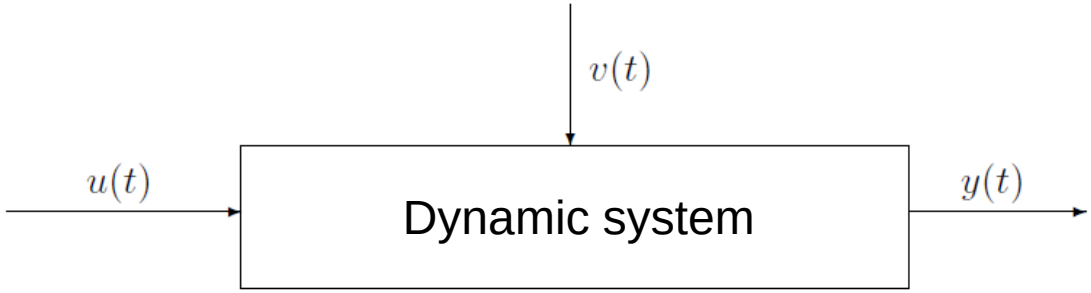


Definition of identification

- *Identification is the experimental determination of the temporal behaviour of a process or system. One uses measured signals and determines the temporal behaviour within a class of mathematical models. The error (respectively deviation) between the real process or system and its mathematical model shall be as small as possible.*
- This definition stems from Zadeh (1962) and introduces three key components:
  - Measured signals
  - Class of mathematical models
  - The error between the real system and its mathematical model
- All three concepts are essential in the classification of the identification procedures as we shall see in the following

Identified system

- Identified system can be illustrated by the following figure



- The system is influenced by the input $u(t)$ and a disturbance $v(t)$
- A modeller can only change (directly or indirectly) the system input while no control over disturbance is possible
- Disturbances can further be classified into:
  - Measured – can be measured directly
  - Unmeasured – their effect can be observed only on the system output $y(t)$

Illustration of important aspects of identification

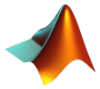
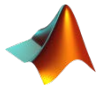
- Identification of a static model 
  - If the wrong model structure is assumed, it seems that every measurement requires its own model
  - If a set of models is a consequence of stochastic process nature, the model that fits a large set of measurements can be selected
- Identification of a real process – a quadcopter 
  - Signal pre-processing – a very important step
  - Determination of model structure – a key step
  - Parameter search method can sometimes be problematic – optimisation can stuck in local minima etc.
- Identification of a noisy static model 
  - Parameter estimate differs from the actual parameter value
  - It is important that the estimate converges to the true value when the number of measurements increases (not always true!)

Illustration of important aspects of identification

- The influence of the noise distribution 
  - If disturbances can be described by the normal distribution, the variance (or the standard deviation) is adequate measure of error
  - In case of other noise distributions, parameter variance is not enough (its distribution is also needed)
  - Due to central limit theorem the distribution of the parameter estimate converges to a normal one
- Noise entering the regressor 
  - Regression equation connects regressor $\boldsymbol{\psi}$, parameter $\boldsymbol{\theta}$ and output y:

$$y = \boldsymbol{\psi}^T \boldsymbol{\theta}$$

  - Noise on the regressor (current in this example) leads to biased estimate – systematic error:
    - Bias size depends on the signal/noise ratio on the regressor
    - Bias does not depend on the output noise (if noises not correlated) – if possible, it is advisable to suitably exchange the output and the regressor

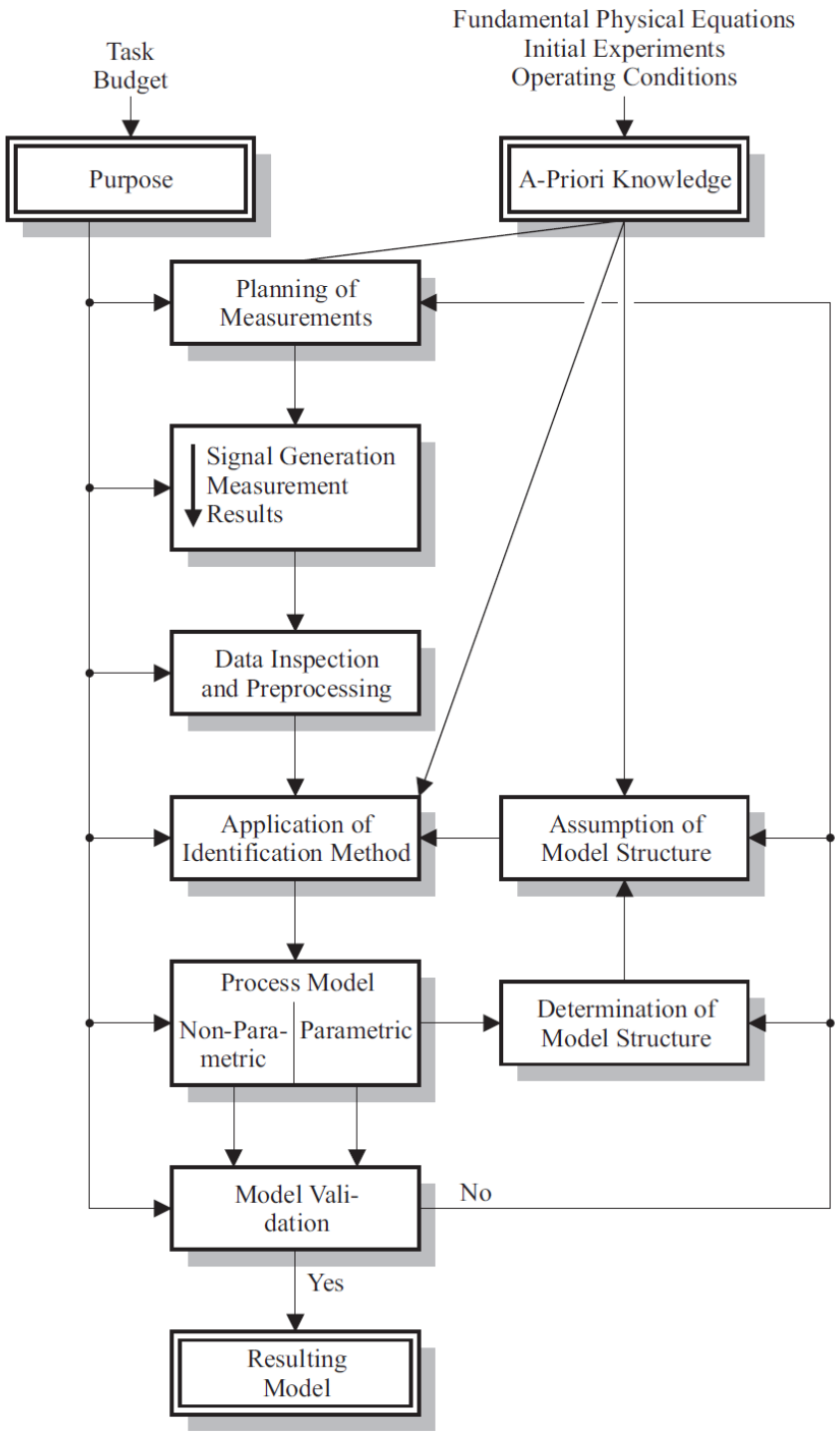
Illustration of important aspects of identification

- Overestimation – the number of estimated parameters is too large:
  - The estimated parameters become correlated
  - The variance of the parameters is higher than the one in the case of a non-overestimation
  - Extrapolation of an identified model is always problematic – in case of overestimation the problem becomes even worse

Practical limitations

- The measured signals are always corrupted by disturbances, noise, drift, outliers or any other type of a „disturbance“
- The available measurement time is always limited
- The maximum allowable change of the input signal, i.e. the test signal height, is always limited – either due to technical reasons or due to the assumption of linearity which is only valid within a certain operating regime
- The maximum allowable change of the output signal may also be limited due to technical reasons or due to the assumption of linear process behaviour
- The shape of input signals is often confined due to an actuator or other hardware properties

Basic sequence of the identification



Steps in the identification process

- **Definition of the model purpose** – essential for model validation
- **Collecting of a-priori knowledge** – encompasses all readily available information about the process to be identified; based on:
  - the knowledge of the process and underlying physical phenomena
  - the experience with the system or similar systems
  - the statistical data about the signals (especially disturbances)

  $\rightarrow$ model structure
- **Planning of measurements / selection of equipment**:
  - input signals (natural – operational, artificial – experimental, the shape, amplitude, frequency spectrum)
  - sampling time and length of experiment
  - open/close loop experiment
  - off-line/on-line identification
  - filtering of disturbances

Steps in the identification process

- **Measurement**
- **Identification algorithm** – determines the identified model based on measurements
- **Model validation** – a test that is done on the measurements not used in the identification algorithm
- If the final model does not fulfil the model purpose, some steps have to be **repeated**

Some remarks:

- Identification model rarely obtained in a single step
- Often some pre-analysis is needed before actual identification is performed
- Identification process is an **iterative** one in general

Classification of identification methods

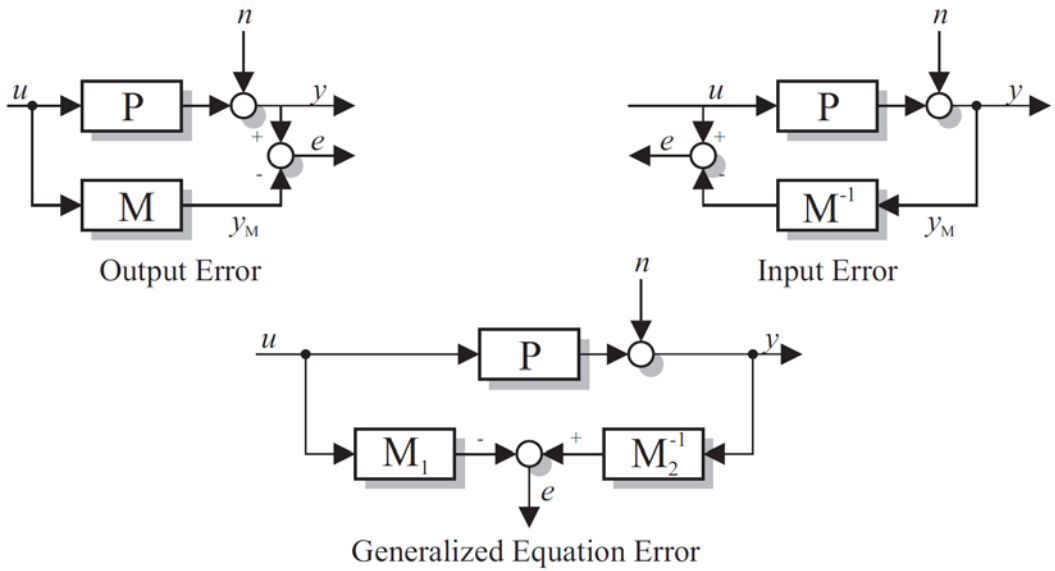
- According to the given definition of identification different identification methods can be classified according to the following criteria:
  - class of mathematical model
  - class of employed test signals
  - calculation of error between process and model
- It has proven practical to also include the following two criteria for classifying the methods further:
  - execution of experiment and evaluation (on-line, off-line)
  - employed algorithm for data processing

Classification of identification methods

- Regarding the class of mathematical model:
  - Parametric identification techniques → models with structure and finite number of parameters (e.g. differential equations, transfer functions)
  - Non-parametric identification techniques → models without specific structure and infinite number of parameters – provide a relation between a certain input and the corresponding response by means of a table or sampled characteristic curve (e.g. impulse response, step response, frequency response – presented in tabular or graphical form)
- Regarding the class of employed input or test signals:
  - According to the continuity of independent variable (time):
    - continuous
    - discrete
  - According to the signals nature:
    - deterministic (analytically describable)
    - stochastic (random)
    - pseudo-stochastic (deterministic – artificially generated, but with properties close to stochastic signals)

Classification of identification methods

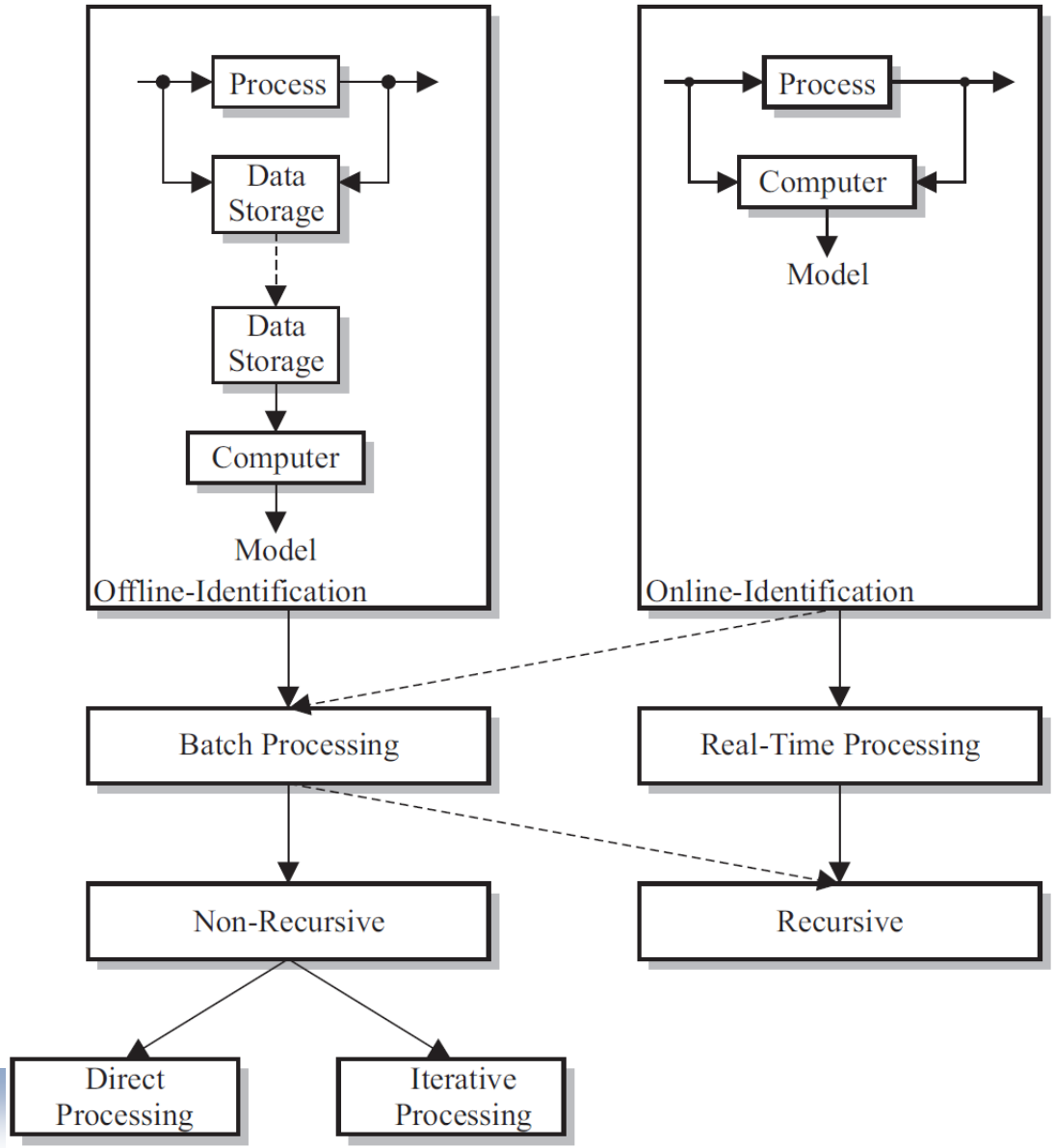
- As a measure for the error between model and process, one can choose between the following errors:
  - Output error (difference between the output of the process and output of the model; both are connected to the same input)
  - Input error (difference between the input of the process and output of the inverse model – its input is connected to the process output)
  - Generalized equation error (model divided into two parts – the rationale behind this approach is to obtain the error model with more favourable properties)



Classification of identification methods

- Two types of *coupling between process and computer*:
  - Offline (indirect coupling) – measured data stored/transferred to the computer for data evaluation and processing
  - Online (direct coupling) – *online identification* performed in parallel to the experiment – the data points processed as they become available
- Classification according to the *type of algorithm* employed:
  - Batch processing – previously stored measurements processed in one shot – typically the case for offline applications
  - Real-time processing – data are processed immediately after they become available – a direct coupling needed
- Another feature is the *processing of the data*:
  - Non-recursive processing – can be direct (in one run) or iterative
  - Recursive processing – updates the model as each measurement becomes available (usually in online mode, but not necessarily)

Classification of identification methods



Identification models

- Parametric models:
  - Continuous models:
    - Differential equations
    - Continuous transfer functions – obtained from differential equations by performing Laplace transform and assuming zero initial conditions
    - State-space description, etc.
  - Discrete models
    - Difference equations
    - Discrete transfer functions – obtained from difference equations by performing z-transform and assuming zero initial conditions
    - State-space description, etc.
- Non-parametric models:
  - Frequency response
  - Step response
  - Impulse response, etc.

Models of continuous systems

- Linear time-invariant (LTI) process with lumped parameters can be described by an ordinary linear differential equation (ODE) with constant coefficients:

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1}\dot{y} + a_n y = b_0 u^{(n)} + b_1 u^{(n-1)} + \cdots + b_{n-1}\dot{u} + b_n u$$

  - $u(t)$ is system input, $y(t)$ its output, b_i and a_i are process parameters
- Performing Laplace transform on ODE and assuming zero initial conditions the process transfer function (TF) is obtained:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1} s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n}$$

- Laplace transform of the output $Y(s)$ is a product between transfer function $G(s)$ and Laplace transform of the input $U(s)$
- Product in frequency domain corresponds to the convolution in time domain:

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau = \int_0^\infty u(\tau)g(t-\tau)d\tau = g(t) * u(t)$$

Models of continuous systems

- Parametric continuous model of a continuous process is characterised by its parameters, e.g., parameters b_i and a_i
- From convolution integral a possible non-parametric model can be inferred: impulse response $g(t) \rightarrow$ defined as the output of a process being excited by an *impulse* (Dirac's delta function) $\delta(t)$
- An alternative non-parametric model – frequency response $G(j\omega)$:
  - Harmonic excitation interpretation $\rightarrow$ If a stable LTI system is excited with a harmonic (sine) signal:

$$u(t) = U_0 \sin(\omega_0 t + \varphi_u)$$

the system responds with a harmonic signal of the same frequency:

$$y(t) = Y_0 \sin(\omega_0 t + \varphi_y)$$

  - Amplitude ratio is defined by amplitude response: $\frac{Y_0}{U_0} = |G(j\omega_0)|$
  - Phase difference is defined by phase response: $\varphi_y - \varphi_u = \measuredangle[G(j\omega_0)]$

Transformations between models of cont. systems

- Impulse response $g(t) \leftrightarrow$ transfer function $G(s)$:

$$G(s) = \mathcal{L}\{g(t)\} \qquad g(t) = \mathcal{L}^{-1}\{G(s)\}$$

- Frequency response $G(j\omega) \leftrightarrow$ transfer function $G(s)$:

$$G(j\omega) = \lim_{s \to j\omega} G(s)$$

- Impulse response $g(t) \leftrightarrow$ frequency response $G(j\omega)$:

$$G(j\omega) = \mathcal{F}\{g(t)\} \qquad g(t) = \mathcal{F}^{-1}\{G(j\omega)\}$$

  - Operator $\mathcal{F}\{\cdot\}$ denotes Fourier transform:

$$X(\omega) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt = \lim_{s \to j\omega} X(s)$$

    - $x(t)$ denotes continuous signal, and $X(\omega)$ its Fourier transform
  - Connection between frequency response of the **system** and Fourier transform of input and output **signals**:

$$G(j\omega) = \frac{Y(\omega)}{U(\omega)}$$

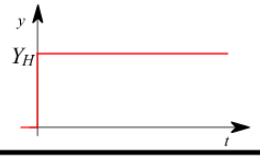
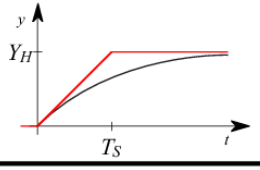
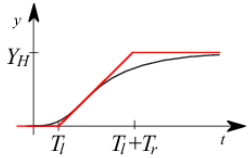
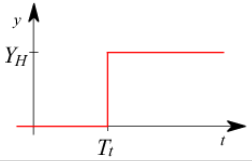
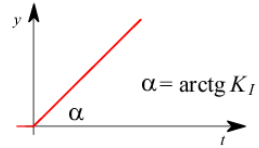
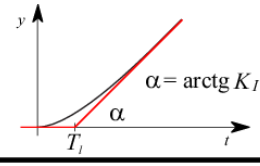
Simple identification methods for cont. systems

- Strejc identification method:
  - Based on the excitation with step test signals
  - Very simple but only suitable for a relatively small class of processes with very humble dynamics but on the other hand such plants occur quite often in practice

- Model tuning methods:
  - Very intuitive
  - Optimisation based $\rightarrow$ computationally demanding
  - Suitable for large class of processes
  - With proper rearrangements also suitable for identification of nonlinear, discrete, time varying etc. systems

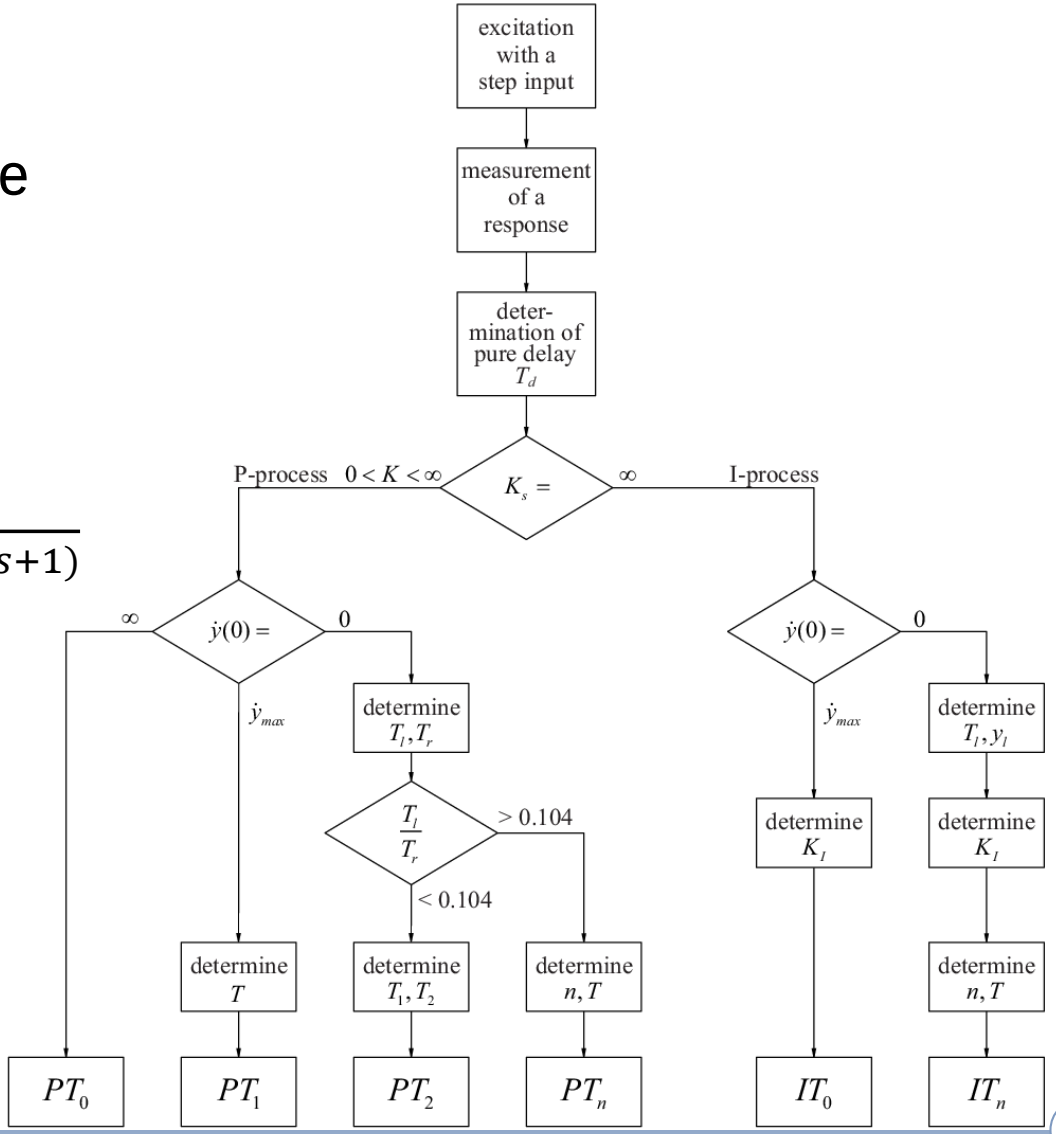
Strejc identification method

- Most often (at least in the area of control design) we deal with proportional (P) and integral (I) processes
- Proportional systems respond with a finite steady-state change when excited by a step input
- Integral systems respond by unbounded response when excited by a step input – asymptote is a polynomial (its order is equal to the number of poles in $s = 0$)

PROCESS	STEP RESPONSE (*step* U_H) (black - actual, red - simplified)	CHARACTERISTIC PARAMETERS		
		K_s	$\dot{Y}_{max}$	*modelled „delay“*
P		$K_S = \frac{Y_H}{U_H}$	∞	0
PT_1		$K_S = \frac{Y_H}{U_H}$	$\frac{Y_H}{T_S}$	0
PT_n (aperiodic) $n \geq 2$		$K_S = \frac{Y_H}{U_H}$	$\frac{Y_H}{T_r}$	T_l
PT_t		$K_S = \frac{Y_H}{U_H}$	∞	T_t
I_1		∞	$K_I U_H$	0
$I_1 T_n$ $n \geq 1$		∞	$K_I U_H$	T_l

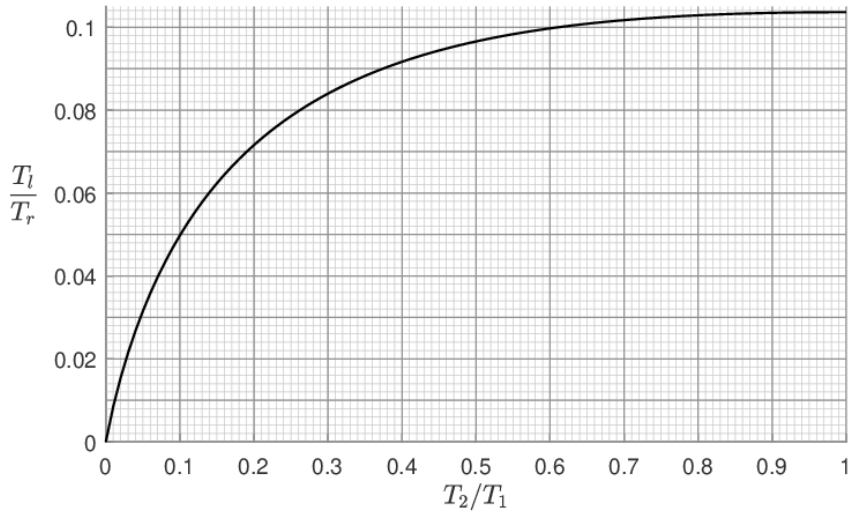
Strejc identification method

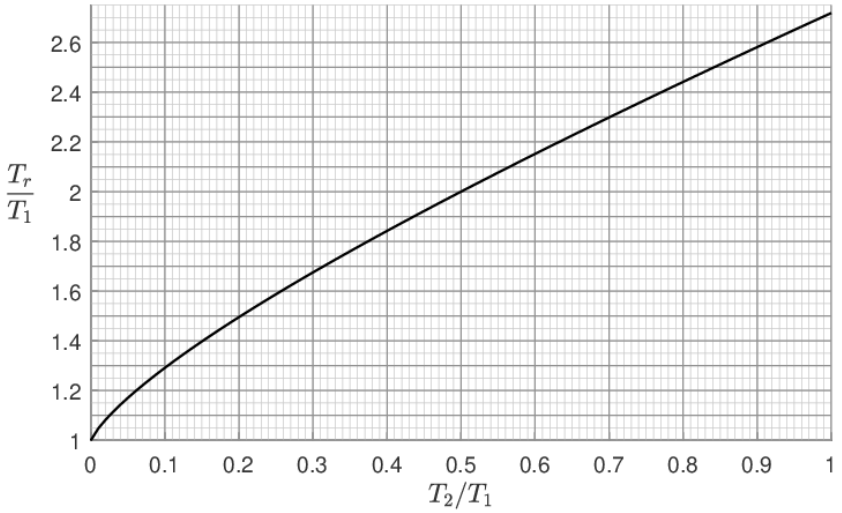
- Flow chart for the determination of simple identification models:
  - PT_0: $G(s) = K_s$
  - PT_1: $G(s) = \frac{K_s}{Ts+1}$
  - PT_2: $G(s) = \frac{K_s}{(T_1 s+1)(T_2 s+1)}$
  - PT_n: $G(s) = \frac{K_s}{(Ts+1)^n}$
  - IT_0: $G(s) = \frac{K_I}{s}$
  - IT_n: $G(s) = \frac{K_I}{s(Ts+1)^n}$



Proportional system of the 2nd order – PT_2

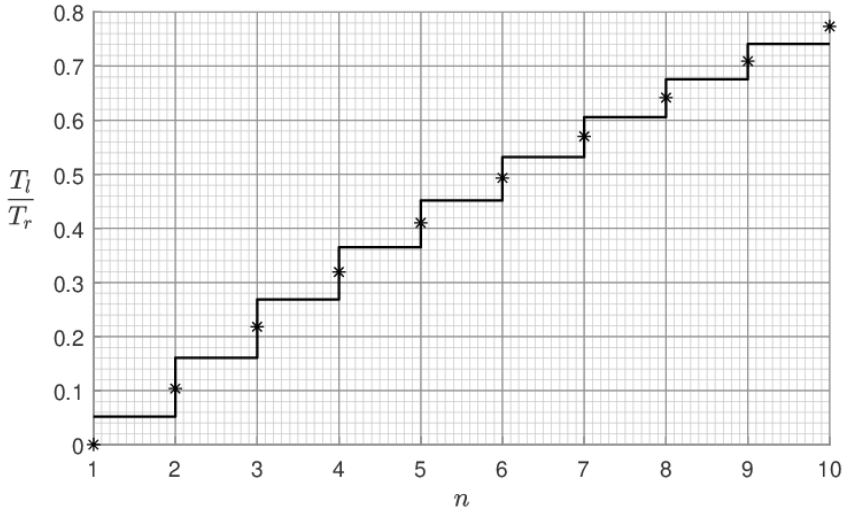
- This model is chosen if
$$\frac{T_l}{T_r} < 0.104$$
- Process model:
$$G(s) = \frac{K_s}{(T_1 s + 1)(T_2 s + 1)}$$
- Measurements needed:
  - T_l
  - T_r
  - $K_s = \frac{\Delta Y}{\Delta U}$
- Procedure:
$$\frac{T_l}{T_r} \rightarrow \frac{T_2}{T_1} \rightarrow \frac{T_r}{T_1} \rightarrow T_1 \rightarrow T_2$$





Proportional system of higher order – PT_n

- This model is chosen if
$$\frac{T_l}{T_r} > 0.104$$
- Process model:
$$G(s) = \frac{K_s}{(Ts+1)^n}$$
- Measurements needed:
  - T_l
  - T_r
  - $K_s = \frac{\Delta Y}{\Delta U}$
- Procedure:
$$\frac{T_l}{T_r} \rightarrow n \rightarrow \frac{T_l}{T} \rightarrow T$$



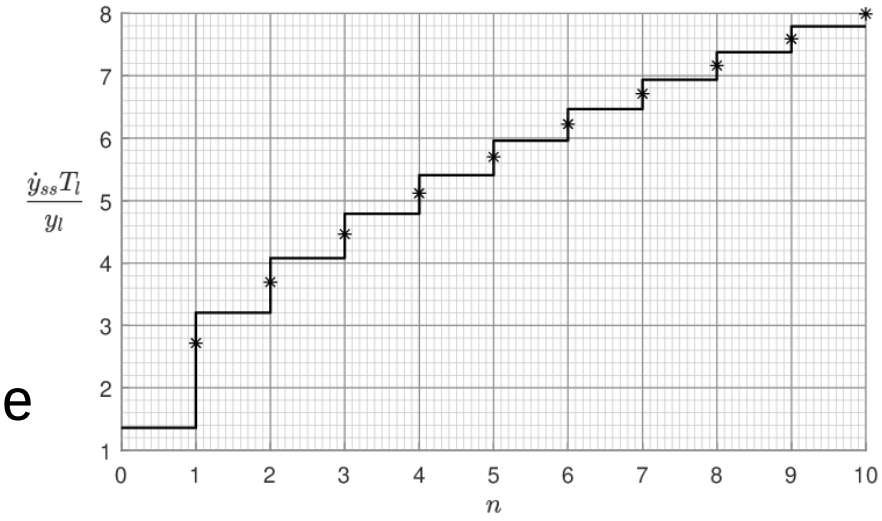
Integral systems (one pole in $s = 0$)

- Process model:

$$G(s) = \frac{K_I}{s(Ts+1)^n}$$

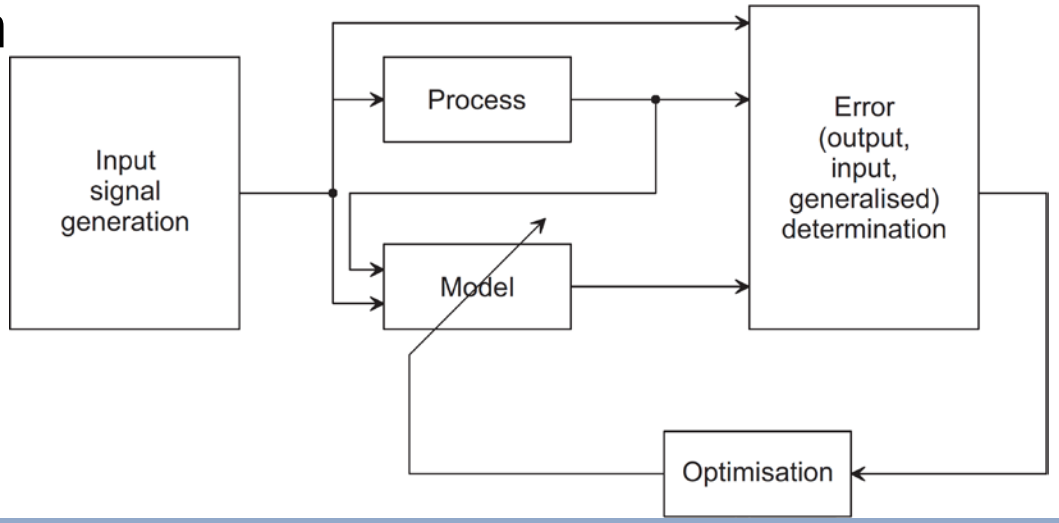
- Response asymptote (after the transient) is a straight line
- Measurements needed:
  - T_l
  - $y_l = y(t)|_{t=T_l}$
  - $K_I = \frac{\dot{y}_{ss}}{\Delta U}$
- Procedure:

$$\frac{\dot{y}_{ss}T_l}{y_l} \rightarrow n \rightarrow \frac{T_l}{T} \rightarrow T$$



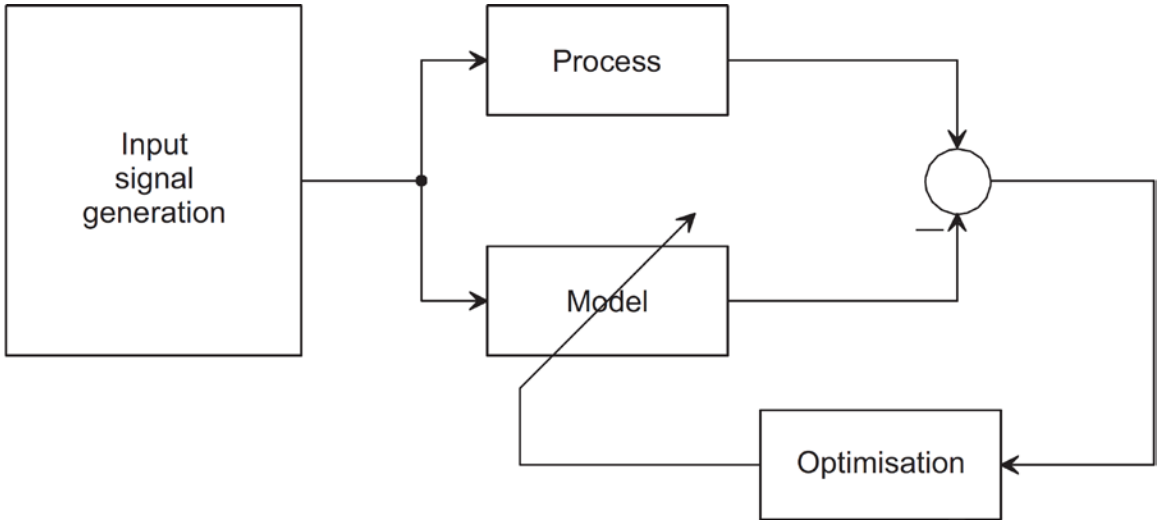
Model tuning methods

- The process and the model are excited with the same signal → output error is the difference between both outputs
- There also exist input error and generalised (equation) error
- Common points: Excitation of the process and the model, **optimisation** of the model (may also include the structure, not only the parameters)
- Useful especially in the off-line version – the signals are stored in some digital form



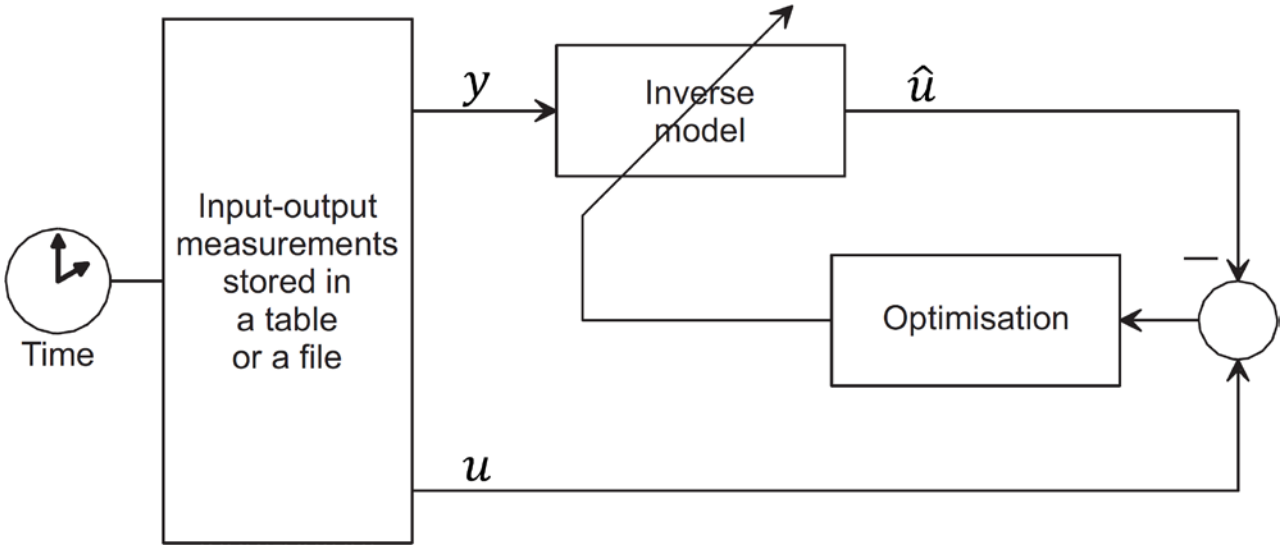
Output-error method

- Process and model are connected in parallel
- Model has to be realised in the recursive form (numerator and denominator of the transfer function are polynomials)
- Approach can be used with continuous and discrete models
- The realisation below takes the signals directly from the process (the version with stored signals also exists)



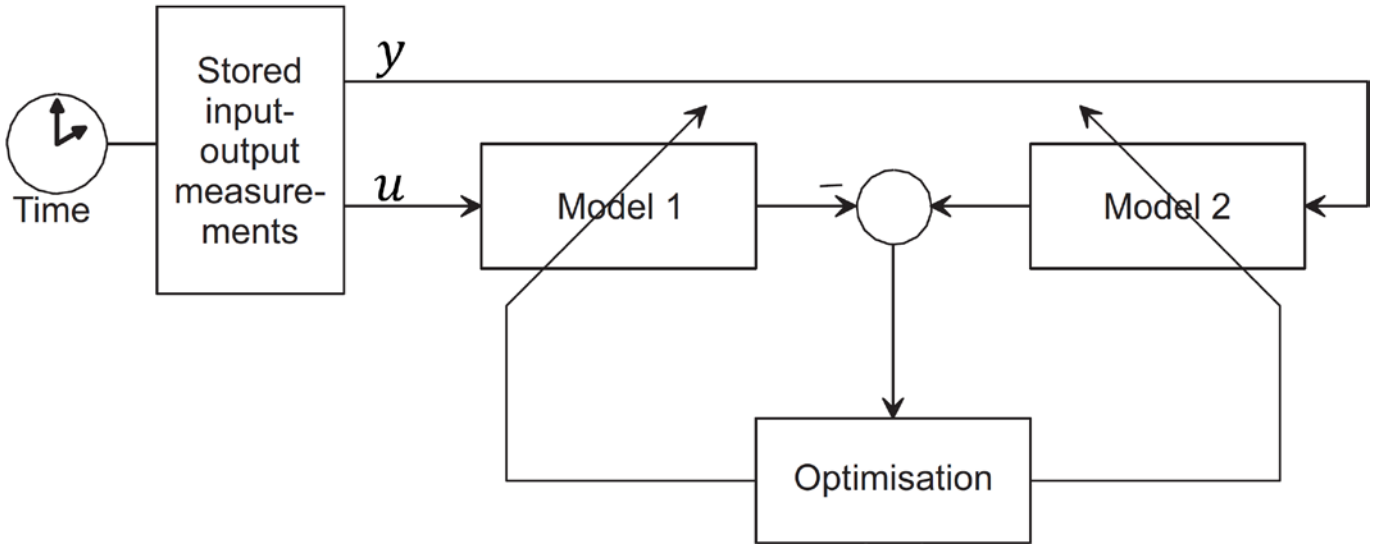
Input-error method

- Realisation with the stored data shown below
- The use limited to systems with causal inverse model, but it is not a problem in off-line mode → the „prediction“ of y (from the data storage) is fed into inverse model instead of actual y thus making the inverse model causal



Generalised-error method

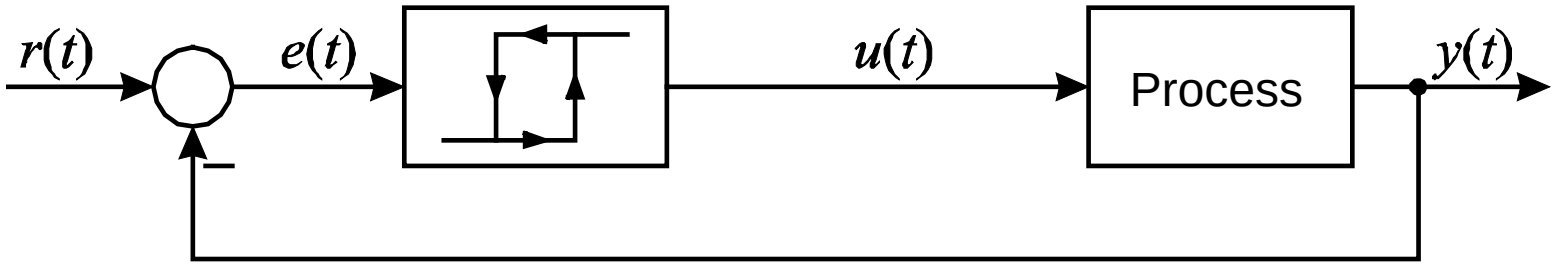
- Generalised-error method is used to make parameter estimation problem linear in parameters
- This is often very useful, e.g., when estimating the transfer function parameters of the discrete systems – model 1 is numerator and model 2 denominator of the transfer function



Simple method for identification of non-parametric models

- Only estimation of the frequency response will be treated:
  - The process is excited by a harmonic signal of a certain frequency
  - The process is linear by assumption → after transient the output becomes harmonic (same frequency, different amplitude and phase)
  - The quotient of amplitudes → amplitude response; phase difference → phase response (both responses are functions of input frequency)
  - Both responses (as a complex function) → frequency response
  - The procedure repeated at all (in practice only interesting) frequencies
  - Measurement of quotient of amplitudes and phase difference:
    - In the past compensation principle
    - Nowadays – mostly correlation methods – noise reduction
  - Two possibilities of excitation generation:
    - Sine-wave signal generator or digital computer: + known frequency
      – run in open-loop
    - Relay feedback

Relay feedback



- In most cases the system starts oscillating around the operating point (OP): + because of feedback the output stays near OP
  – frequency needs to be estimated – not very easy – period of oscillation not multiple of sampling time
- Signal $u(t)$ is a series of pulses, $y(t)$ close to a sine (u and y have opposite phase)
- Åstrom and Hägglund patented the approach for PID tuning
- System analysis only possible with describing functions; here we show a simulation example

Correlation functions

- Measured data are almost always corrupted with noise → theory of stochastic signals needed
- If (stochastic) signals are mutually dependent, the dependency can be evaluated by correlation functions:
  - Auto-correlation function evaluates the dependency between the signal and its time shift:

$$\phi_{uu}(\tau) = E\{u(t)u(t+\tau)\}$$

  - According to the *ergodic hypothesis*, one can obtain the same statistical information (which one gets from ensemble averaging) also from averaging a single sample function over time if infinitely long intervals of time are considered: $E\{\cdot\} \rightarrow \int_{-\infty}^{\infty}$

$$\phi_{uu}(\tau) = E\{u(t)u(t+\tau)\} = \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} u(t)u(t+\tau)dt$$

    - T_i is the integration interval

Correlation functions

- The same holds for cross-correlation functions:

$$\phi_{uy}(\tau) = E\{u(t)y(t+\tau)\} = \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} u(t)y(t+\tau)dt$$

- If signals are only measured from $t = 0$, a slight modification of integral bounds is introduced:

$$\phi_{uu}(\tau) = E\{u(t)u(t+\tau)\} = \lim_{T_i \to \infty} \frac{1}{T_i} \int_{0}^{T_i} u(t)u(t+\tau)dt$$

$$\phi_{uy}(\tau) = E\{u(t)y(t+\tau)\} = \lim_{T_i \to \infty} \frac{1}{T_i} \int_{0}^{T_i} u(t)y(t+\tau)dt$$

Correlation functions

- The link between auto- and cross-correlation:

$$\phi_{uy}(\tau) = \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} u(t) \int_0^{\infty} g(\sigma) u(t + \tau - \sigma) d\sigma dt$$

$$= \int_0^{\infty} g(\sigma) \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} u(t) u(t + \tau - \sigma) dt d\sigma$$

$$= \int_0^{\infty} g(\sigma) \phi_{uu}(\tau - \sigma) d\sigma = \int_0^{\infty} g(t) \phi_{uu}(\tau - t) dt$$

- Cross-correlation between input and output signal is a convolution of auto-correlation and the impulse response:

$$\phi_{uy}(\tau) = g(\tau) * \phi_{uu}(\tau) = \phi_{uu}(\tau) * g(\tau)$$

$\rightarrow$ impulse response $g(\tau)$ can be obtained as a deconvolution between the cross-correlation $\phi_{uy}(\tau)$ and auto-correlation $\phi_{uu}(\tau)$

Correlation functions and spectra

- Similar relation holds for $\phi_{yu}(\tau)$ and $\phi_{yy}(\tau)$:

$$\phi_{yy}(\tau) = \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} y(t) \int_0^{\infty} g(\sigma) u(t + \tau - \sigma) d\sigma dt$$

$$= \int_0^{\infty} g(t) \phi_{yu}(\tau - t) dt$$

- Note that $\phi_{yu}(\tau)$ and $\phi_{uy}(\tau)$ are different but closely related cross-correlations:

$$\phi_{yu}(\tau) = \phi_{uy}(-\tau)$$

- Fourier transforms of a correlation function is **power spectral density** (or simply **power spectrum**):

$$\Phi_{uu}(\omega) = \mathcal{F}\{\phi_{uu}(\tau)\}$$
$$\Phi_{uy}(\omega) = \mathcal{F}\{\phi_{uy}(\tau)\}$$
$$\Phi_{yy}(\omega) = \mathcal{F}\{\phi_{yy}(\tau)\}$$

Correlation functions and spectra

- Relations between power spectral densities (PSD):

$$\Phi_{uy}(\omega) = \int_{-\infty}^{\infty} \phi_{uy}(\tau) e^{-j\omega\tau} d\tau$$

$$= \int_{-\infty}^{\infty} \int_{0}^{\infty} g(t)\phi_{uu}(\tau - t)dt \; e^{-j\omega\tau} d\tau$$

$$= \int_{0}^{\infty} g(t)e^{-j\omega t} \int_{-\infty}^{\infty} \phi_{uu}(\tau - t)e^{-j\omega(\tau - t)} d\tau dt = G(j\omega)\Phi_{uu}(\omega)$$

- Similarly:

$$\Phi_{yy}(\omega) = G(j\omega)\Phi_{yu}(\omega)$$

- Relation between $\Phi_{uu}(\omega)$ and $\Phi_{yy}(\omega)$:

$$\Phi_{yy}(\omega) = G(j\omega)\Phi_{yu}(\omega) = G(j\omega)\Phi_{uy}(-\omega) =$$
$$= G(j\omega)G(-j\omega)\Phi_{uu}(-\omega) = |G(j\omega)|^2 \Phi_{uu}(\omega)$$

  - We took into account: $\Phi_{yu}(\omega) = \Phi_{uy}(-\omega)$ in $\Phi_{uu}(-\omega) = \Phi_{uu}(\omega)$

Correlation functions and spectra

- Relation between Fourier transforms of two signals and their cross-power spectrum:

$$\Phi_{uy}(\omega) = \int_{-\infty}^{\infty} \phi_{uy}(\tau) e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} u(t) y(t+\tau) dt \; e^{-j\omega\tau} d\tau$$

$$= \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} u(t) e^{j\omega t} \int_{-\infty}^{\infty} y(t+\tau) e^{-j\omega(t+\tau)} d\tau dt$$

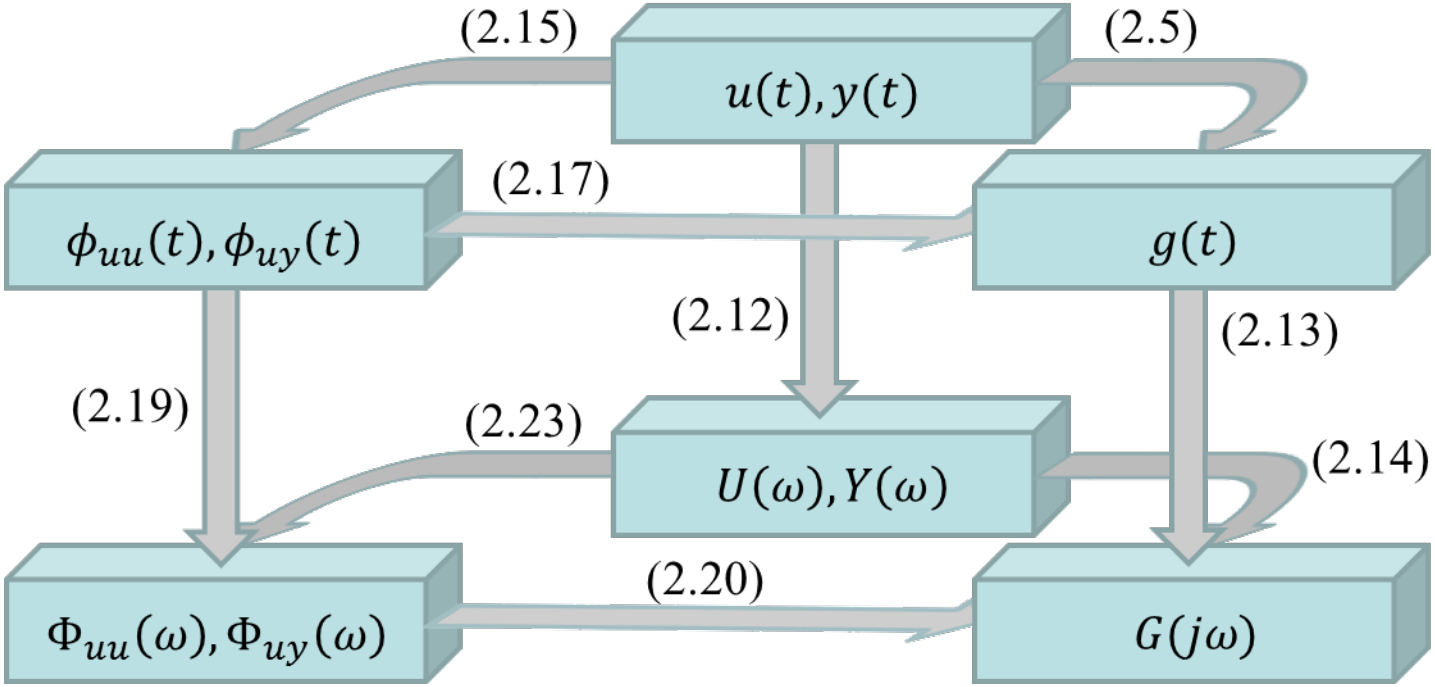
$$= \lim_{T_i \to \infty} \frac{1}{T_i} \int_{-\frac{T_i}{2}}^{\frac{T_i}{2}} u(t) e^{j\omega t} Y(\omega) dt = \lim_{T_i \to \infty} \frac{1}{T_i} U(-\omega) Y(\omega)$$

- In case we are dealing with only one signal:

$$\Phi_{uu}(\omega) = \lim_{T_i \to \infty} \frac{1}{T_i} U(-\omega) U(\omega) = \lim_{T_i \to \infty} \frac{1}{T_i} |U(\omega)|^2$$

Correlation functions and spectra

- Graphical representation between input and output signals, their correlations, power spectra and non-parametric models:



- References to equations in the textbook: D. Matko, Identifikacije Založba FE, Ljubljana, 1999.

Models of discrete systems

- Linear time-invariant (LTI) process with lumped parameters can be described by an ordinary linear difference equation with constant coefficients:

$$y(k) + a_1 y(k-1) + \cdots + a_{n-1} y(k-n+1) + a_n y(k-n) =$$
$$= b_0 u(k) + b_1 u(k-1) + \cdots + b_{n-1} u(k-n+1) + b_n u(k-n)$$

  - $u(k)$ is process input, $y(k)$ is its output, b_i and a_i are its parameters
- Performing z transform on this difference equation and assuming zero initial conditions the process transfer function (TF) is obtained:

$$G(z) = \frac{Y(z)}{U(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_{n-1} z^{-n+1} + b_n z^{-n}}{1 + a_1 z^{-1} + \cdots + a_{n-1} z^{-n+1} + a_n z^{-n}}$$

- z-transform of the output $Y(z)$ is a product between TF and $U(z)$
- Product in frequency domain corresponds to the convolution in time domain:

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

Models of discrete systems

- Parametric discrete model of a discrete process is characterised by its parameters, e.g., parameters b_i and a_i
- From convolution a possible non-parametric model can be inferred:
  - impulse response $g(k) \rightarrow$ the response to *impulse* $\delta(k) = \begin{cases} 1, k = 0 \\ 0, k \neq 0 \end{cases}$
    - discrete impulse $\delta(k) \neq \delta(t)$
- An alternative non-parametric model – frequency response $G\left(e^{j\omega T}\right)$:
  - Harmonic excitation interpretation $\rightarrow$ If a stable LTI system is excited with a harmonic (sine) signal:

$$u(k) = U_0 \sin(\omega_0 kT + \varphi_u)$$

the system responds with a harmonic signal of the same frequency:

$$y(k) = Y_0 \sin(\omega_0 kT + \varphi_y)$$

  - Amplitude ratio is defined by amplitude response: $\frac{Y_0}{U_0} = \left|G\left(e^{j\omega_0 T}\right)\right|$
  - Phase difference is defined by phase response: $\varphi_y - \varphi_u = \measuredangle[G(e^{j\omega_0 T})]$

Transformations between models of disc. systems

- Impulse response $g(k) \leftrightarrow$ transfer function $G(z)$:

$$G(z) = \mathcal{Z}\{g(k)\} \qquad g(k) = \mathcal{Z}^{-1}\{G(z)\}$$

- Frequency response $G\left(e^{j\omega T}\right) \leftrightarrow$ transfer function $G(z)$:

$$G\left(e^{j\omega T}\right) = \lim_{z \to e^{j\omega T}} G(z)$$

- Impulse response $g(k) \leftrightarrow$ frequency response $G\left(e^{j\omega T}\right)$:
  - Usually discrete Fourier transform (or its fast version FFT) is used → spectrum is discrete as shown in the following
- Relations between correlations and spectra are straight-forward extension of the continuous ones:
  - $\phi_{uy}(\tau) = \sum_{k=0}^{\infty} g(k)\phi_{uu}(\tau - k) = g(\tau) * \phi_{uu}(\tau)$
  - $\Phi_{uy}(\omega) = G(e^{j\omega T})\Phi_{uu}(\omega)$
  - $\Phi_{yy}(\omega) = |G(e^{j\omega T})|^2 \Phi_{uu}(\omega)$

Relation between continuous and discrete signals

- Real signals (measurements) usually continuous and aperiodic
- Processing on digital computers involves Fast Fourier Transform (FFT) which assumes discrete and periodic signals
- Denotations:
  - Continuous time (CT) signal with small letter, e.g. $x(t)$
  - Fourier transform of CT signal with capital letter, e.g. $X(\omega)$
  - Sampled signal with a star, e.g. $x^*(t)$, $X^*(\omega)$
  - Periodic signals with superscript p: CT signal $x^p(t)$, its Fourier transform / coefficient of Fourier series $X^p(m\frac{2\pi}{T})$
  - Discrete time (DT) and periodic signals with superscript d: DT signal $x^d(k)$, its discrete Fourier transform $X^d(m)$
  - Sampling time of sampled signals is T, the period is t_p
  - Frequency quant (for FFT) is F, the period of frequency signals is the same as sampling frequency f_s

Relation continuous-discrete in time domain

- Assume that the original continuous time and aperiodic signal:
  - only exists (is different from 0) in the interval $[0, t_p)$
  - outside of this interval the signal is assumed 0

$$x(t) = 0 \qquad t < 0 \text{ or } t \geq t_p = NT$$

  - close to practical case – measurement interval is of finite length
- Sampled signal obtained by multiplying original signal $x(t)$ with a series of delta-impulses:

$$x^*(t) = x(t) \sum_{k=-\infty}^{\infty} \delta(t - kT)$$

- Due to the fact that $x(t)$ different from 0 only if $t \in [0, NT)$, the summation bounds can change to 0 and $N - 1$, respectively:

$$x^*(t) = x(t) \sum_{k=0}^{N-1} \delta(t - kT) = \sum_{k=0}^{N-1} x(kT)\delta(t - kT)$$

Relation continuous-discrete in time domain

- On digital computers only sampled signals can be used (in time and frequency domain)
- Sampling in one domain results in periodicity in the other and therefore all the signals are periodic
- Assume that the periodic signal is obtained by making shifted copies ($\pm t_p, \pm 2t_p, \ldots$) of original signals:

$$x^p\left(t \pm l t_p\right) = x(t) \quad t \in \left[0, t_p\right), l = 0,1,2, \ldots$$

- Digital signal $x^d(k)$ is obtained by sampling the original $x(t)$; k is an integer:

$$x^d(k) = x(t)\Big|_{t=kT} \qquad k \in [0, N-1]$$

- Only periodic digital signals will be treated:

$$x^d(k) = x^p(kT)$$

Relation continuous-discrete in frequency domain

- Fourier transform of a continuous signal:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt = \int_{0}^{t_p} x(t)e^{-j\omega t}dt$$

  - $X(\omega)$ continuous and aperiodic
- In case of periodic signals, Fourier series is used instead of Fourier transform:
  - The spectrum becomes discrete – only includes frequency 0 (DC component) and frequencies $\pm\frac{1}{t_p}$, $\pm\frac{2}{t_p}$, ... (basic and higher harmonics)
- Fourier series coefficients:

$$X^p(m\tfrac{2\pi}{t_p}) = \frac{1}{t_p}\int_{0}^{t_p} x^p(t)e^{-jm\frac{2\pi}{t_p}t}dt = \frac{1}{t_p}\int_{0}^{t_p} x(t)e^{-jm\frac{2\pi}{t_p}t}dt$$

- Relation between $X(\cdot)$ and $X^p(\cdot)$: $X(\omega)|_{\omega=m\frac{2\pi}{t_p}} = t_p X^p(m\tfrac{2\pi}{t_p})$

Relation continuous-discrete in frequency domain

- Fourier transform of digital signals (sampled and periodic) is calculated by means of discrete Fourier transform (computationally efficient algorithms is called FFT) :

$$X^d(m) = \sum_{k=0}^{N-1} x^d(k) e^{-j\frac{2\pi}{N}mk}$$

$$x^d(k) = \frac{1}{N} \sum_{m=0}^{N-1} X^d(m) e^{j\frac{2\pi}{N}mk}$$

- Fourier series coefficients and Fourier transform samples directly related to coefficients of DFT/FFT:

$$X^p\left(m\frac{2\pi}{t_p}\right) = \frac{1}{N}X^d(m) \qquad m < \frac{N}{2}$$

$$X(\omega)|_{\omega=m\frac{2\pi}{t_p}} = t_p X^p\left(m\frac{2\pi}{t_p}\right) = TX^d(m) \qquad \omega < \frac{\pi}{T}$$

Relation continuous-discrete in frequency domain

- Discrete cross-correlation defined as:

$$\phi_{xy}^{d}(\tau_i) = \frac{1}{N}\sum_{k=0}^{N-1} x^{d}(k)y^{d}(k+\tau_i)$$

  - τ_i is an integer and N the period of $\phi_{xy}^{d}(\cdot)$
- Cross-correlation of sampled periodic signals:

$$\phi_{xy}^{*p}(\tau_i T) = \begin{cases} \frac{1}{T}\phi_{xy}^{d}(\tau_i) & \tau_i = 0, \pm 1, \pm 2, \ldots \\ 0 & \tau_i \text{ not an integer} \end{cases}$$

- Relation between cross-correlations of periodic and aperiodic signals:

$$\phi_{xy}^{p}(\tau) = \frac{1}{t_p}\phi_{xy}(\tau)$$

Relation continuous-discrete in frequency domain

- Power spectrum:

$$\Phi_{xy}^{p}\left(m\frac{2\pi}{t_p}\right)=\frac{1}{N}\Phi_{xy}^{d}(m) \qquad m<\frac{N}{2}$$

- Power spectral density:

$$\Phi_{xy}(\omega)\Big|_{\omega=m\frac{2\pi}{t_p}}=t_p\Phi_{xy}^{p}\left(m\frac{2\pi}{t_p}\right) \qquad \omega<\frac{\pi}{T}$$

$$=T\Phi_{xy}^{d}(m)$$

- Power spectral density is obtained by distributing the power spectrum over the interval between two samples of power spectrum, namely dividing the spectrum by the frequency quant $F=\frac{1}{t_p}$ or alternatively multiplying it by t_p

The concept of disturbances

- Disturbances are very important factor in identification
- This term includes all the measurement errors, effects of external signals, unmodelled dynamics etc.
- If linearity of a process is assumed, all these disturbances except input measurement error can be collected in one term $n(t)$ which is added to the undisturbed output y_0:

$$y(t) = y_0(t) + n(t)$$

- Disturbances are unpredictable by assumption but still can be classified into some typical categories:
  - high-frequency quasi-stationary stochastic signals (noise)
  - low-frequency non-stationary disturbances, e.g. drift or step (often as a result of the wrong estimation of the operating point)
  - harmonic disturbances
  - unknown disturbances such as spikes, outliers etc.

Noise

- In identification noise can take one of the two roles:
  - disturbances that can include measurement errors, immeasurable influences of some other variables etc.
  - excitation signals with rich frequency content
- To describe signals one must distinguish between:
  - (Amplitude) distribution
    - It gives probability that the signal amplitude is inside of the certain interval
  - Frequency spectrum or frequency content
    - It gives information about the correlation between signal values in consecutive time instants – this dependency is closely related to the frequency content of the signal $\rightarrow$ spectrum
    - *White noise* is a special type of noise that includes all the frequencies with equal shares (like white light); the consequence is that two signal values taken arbitrary close in time are statistically independent (prediction not possible)

Distribution of stochastic signals

- Defined by the *probability density function* (*PDF*) $p(x)$, for which $\int_{-\infty}^{\infty} p(x)dx = 1$ is always fulfilled. Some common PDFs:
  - Normal or Gaussian distribution:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

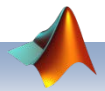
    - $\bar{x}$ is a mean value of the signal and σ its standard deviation
    - the sum of two signals with Gaussian distribution is Gaussian
  - Uniform distribution:

$$p(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

    - PDF is constant in the $[a, b]$ interval
  - Binary distribution:

$$p(x) = \tfrac{1}{2}[\delta(x) + \delta(x-1)]$$

    - discrete distribution $\rightarrow$ PDF composed of δ impulses



White noise

- Power spectral density of a white noise is constant by definition:

$$\Phi_{vv}(\omega) = \Phi_0$$

- White noise $v(t)$ is a signal with statistically independent values at different time instants
- Autocorrelation is therefore a δ impulse
- Usually it is assumed that the white noise mean value is 0 $\rightarrow$ its autocorrelation can be described by:

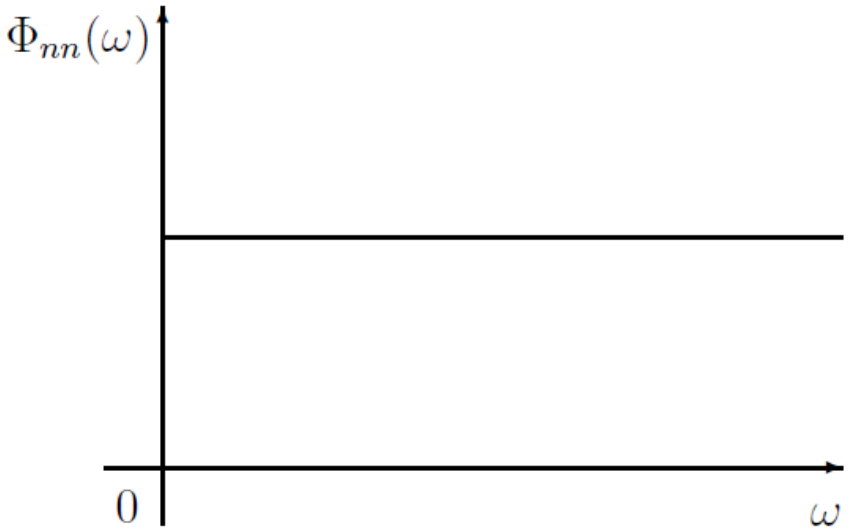
$$\phi_{vv}(\tau) = \Phi_0 \delta(\tau)$$

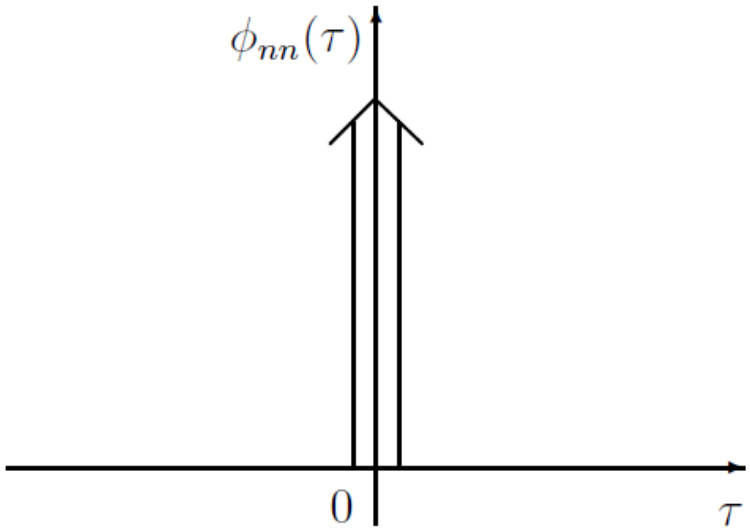
  - Φ_0 is power spectral density of white noise
- White noise seems as a perfect candidate for system excitation in the context of system identification, BUT:

$$\lim_{t_p \to \infty} \frac{1}{t_p} \int_{-\frac{t_p}{2}}^{\frac{t_p}{2}} v^2(t)dt = \phi_{vv}(0) = \infty \quad \rightarrow \textbf{infinite power}$$

White noise

- Power spectral density (PSD) and signal autocorrelation



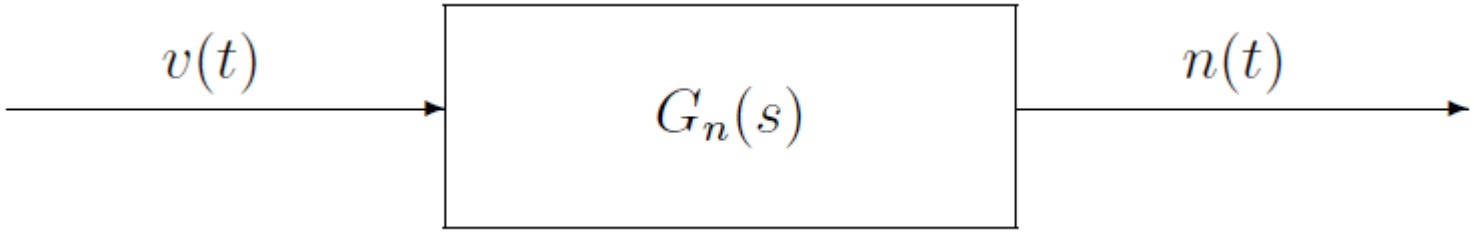


Coloured noise

- Coloured noise does not include all frequencies or at least its PSD is not constant
- PSD of a signal can be changed by applying a filter:

$$\Phi_{yy}(\omega) = |G(j\omega)|^2 \Phi_{uu}(\omega)$$

- PSD is called *rational* if it can be obtained by filtering white noise through a filter (given by a rational, i.e. „normal“, transfer function); the filter $G_n(s)$ is called a noise filter



- The above coloured noise generation makes analysis easier
- The actual generation is not possible in this way!

Coloured noise

- If a noise at the filter input is white, PSD of a coloured noise at the noise filter output becomes:

$$\Phi_{nn}(\omega) = |G_n(j\omega)|^2 \Phi_0$$

- If $G_n(s)$ is stable, the coloured noise is *stationary*
- The coloured noise with an arbitrary PSD can be obtained by filtering white noise through a stable and minimum phase noise filter:
  - This is the result of the absolute value of $|G_n(j\omega)|$ in the above equation $\rightarrow$ the influence of a pole (or a zero) in the left half-plane is the same as in the case of a pole (or a zero) in the right half-plane
  - A pole (or a zero) of a noise filter can always be moved symmetrically over the imaginary axis without affecting the PSD at the output
  - In the case of discrete signals unit circle is used instead of imag. axis
- BUT: the filter order can become extremely high („strange“ PSD)!

Coloured noise

- *Broad-band noise* is a special case of a coloured noise where a low-pass filter of the 1st order is used (high cut-off frequency ω_g):

$$G_n(s) = \frac{1}{1 + sT_g}$$

  - $T_g = \frac{1}{\omega_g}$ is a time constant of the low pass filter

- Power spectral density of a broad-band noise:

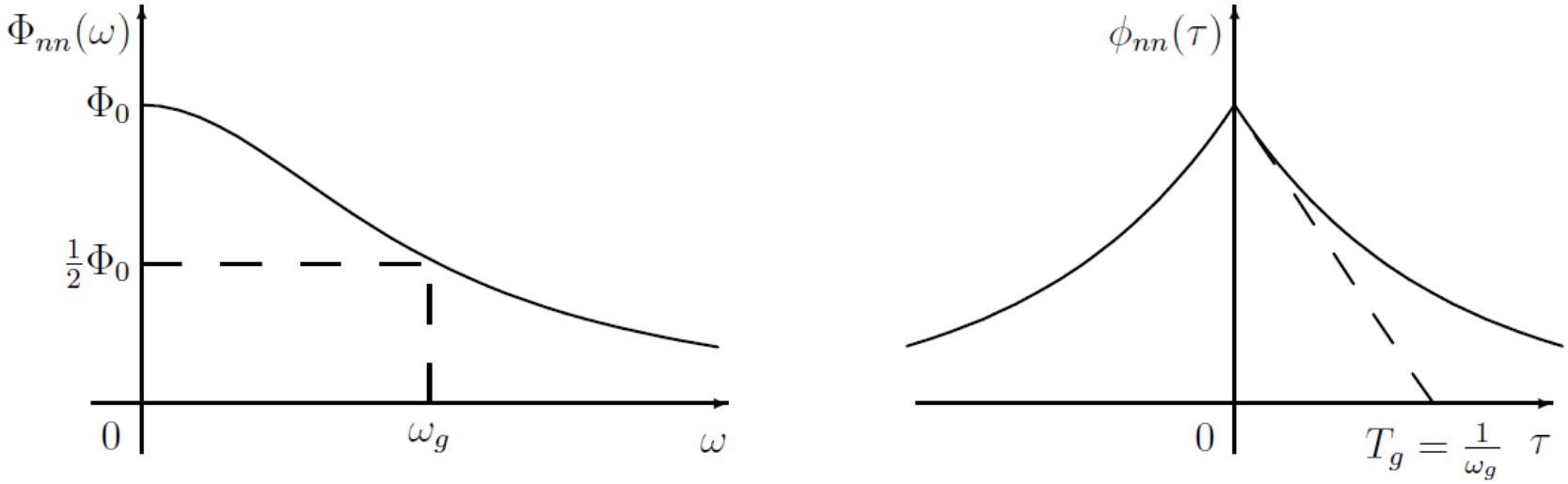
$$\Phi_{nn}(\omega) = |G_n(j\omega)|^2 \Phi_0 = \frac{\Phi_0}{1 + T_g^2 \omega^2} = \frac{\Phi_0}{1 + \left(\frac{\omega}{\omega_g}\right)^2}$$

- Auto-correlation of the broad-band noise:

$$\phi_{nn}(\tau) = \frac{\Phi_0 e^{-\left|\frac{\tau}{T_g}\right|}}{2T_g} = \frac{\Phi_0 e^{-|\tau|\omega_g}}{2} \omega_g$$

Coloured noise

- Power spectral density (PSD) and autocorrelation function of the broad-band noise:



- Mean power (or mean square or variance): $\phi_{nn}(0) = \frac{\Phi_0}{2T_g} = \frac{\Phi_0 \omega_g}{2}$
  - At small T_g (or equivalently large ω_g) broad-band noise approaches white noise:
    - Autocorrelation tends to δ impulse (with area Φ_0)
    - PSD almost constant on a large frequency interval

Random Binary Signal (RBS)

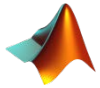
- *Random Binary Signal* is a continuous-time random binary signal:
  - The signal amplitude can take two states: $+a$ and $-a$
  - The change of signal value is random and is not synchronised with anything – infinitely fast changes are theoretically possible
- Also called „random telegraph signal“
- If compared with a random signal with continuous amplitude distribution RBS has some advantages:
  - RBS has the highest PSD among all the signals with bounded amplitude (good excitation necessary for high quality of the model – „good“ can also mean high PSD, among other)
  - Simple generation when hardware is limited, e.g. relay based
  - Simple calculation of the cross correlation (just multiplication of the output signal with $+a$ or $-a$)

Autocorrelation function of RBS

- We assume that the probability of n sign changes ($a \to -a$ or $-a \to a$) in a given period of time Δt is Poisson-distributed with:

$$P(n) = \frac{(\mu \Delta t)^n}{n!} e^{-\mu \Delta t}$$ (μ is the average number of sign changes over a time unit)

- The probabilities of sign changes in Δt are therefore:
  - 0 sign changes $\to P(0) = e^{-\mu \Delta t}$
  - 1 sign change $\to P(1) = \mu \Delta t e^{-\mu \Delta t}$
  - 2 sign changes $\to P(2) = \frac{(\mu \Delta t)^2}{2!} e^{-\mu \Delta t}$ ...



- Product $u(t)u(t+\tau)$ for RBS can be either $+a^2$ ali $-a^2$ ($u(t)$ and $u(t+\tau)$ can either be of the same or different sign)
- $E\{u(t)u(t+\tau)\} = \begin{cases} +a^2 & \text{even number of sign changes from } t \text{ to } t+\tau \\ -a^2 & \text{odd number of sign changes from } t \text{ to } t+\tau \end{cases}$
- $E\{u(t)u(t+\tau)\} = +a^2$ at $\tau = 0$

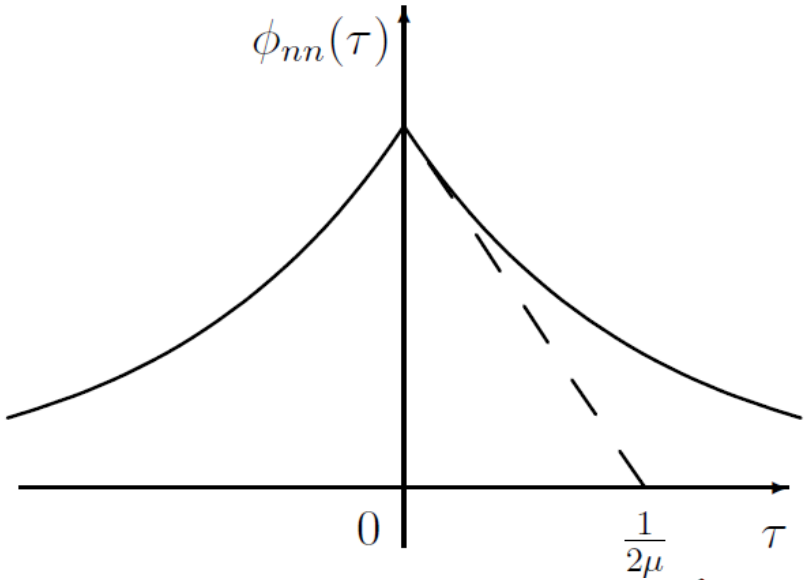
Autocorrelation function of RBS

- Sign changes are random, but mathematical expectation of $u(t)u(t+\tau)$ can be obtained based on the probability of the events:

$$E\{u(t)u(t+\tau)\} = +a^2[P(0)+P(2)+\cdots] - a^2[P(1)+P(3)+\cdots]$$

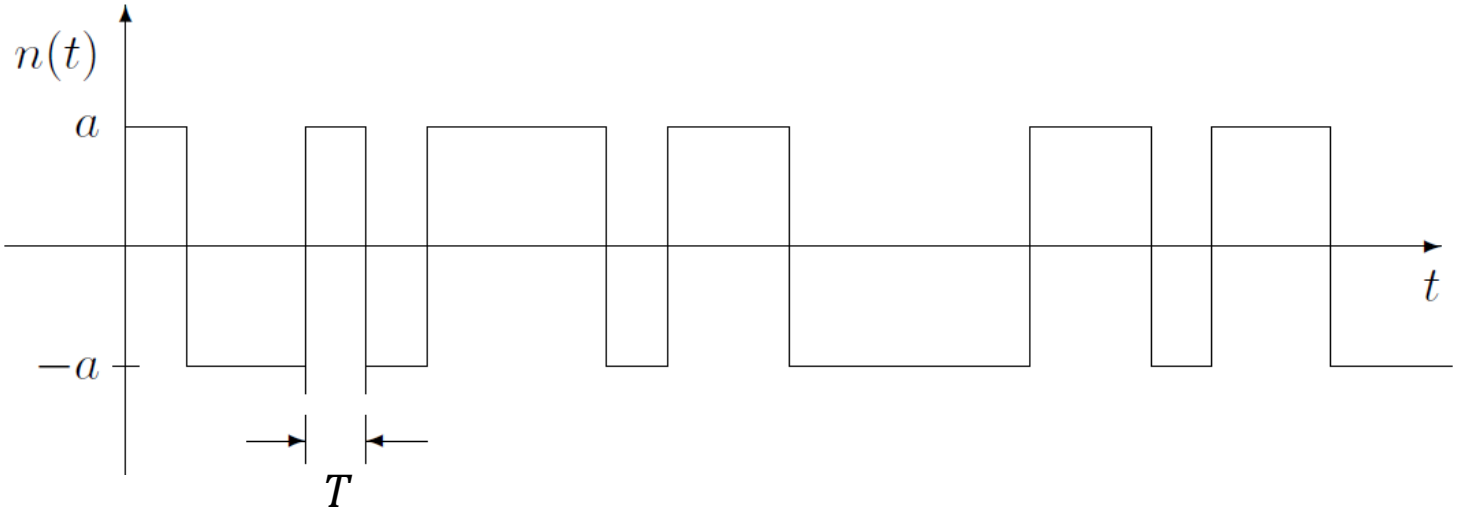
$$= a^2 e^{-\mu|\tau|}\left[1 - \frac{\mu|\tau|}{1!} + \frac{(\mu|\tau|)^2}{2!} - \frac{(\mu|\tau|)^3}{3!} + \cdots\right] = a^2 e^{-2\mu|\tau|}$$

- The same as in the case of broad-band noise if:
  - $a^2 = \frac{\Phi_0}{2}\omega_g$
  - $\mu = \frac{\omega_g}{2}$
- Main problem: the times between sign changes can be arbitrary short (practical implementation not possible!)



Discrete Random Binary Signal (DRBS)

- *Discrete Random Binary Signal* (DRBS):
  - an approximation of RBS – makes implementation easier
  - solves the problem of possibly extremely fast sign changes in RBS
  - sign changes can only occur in discrete moments $t = kT, k = 1,2,3 \ldots$ (T is the length of the sampling time, also referred to as cycle time)
- Typical shape:



Autocorrelation function of DRBS

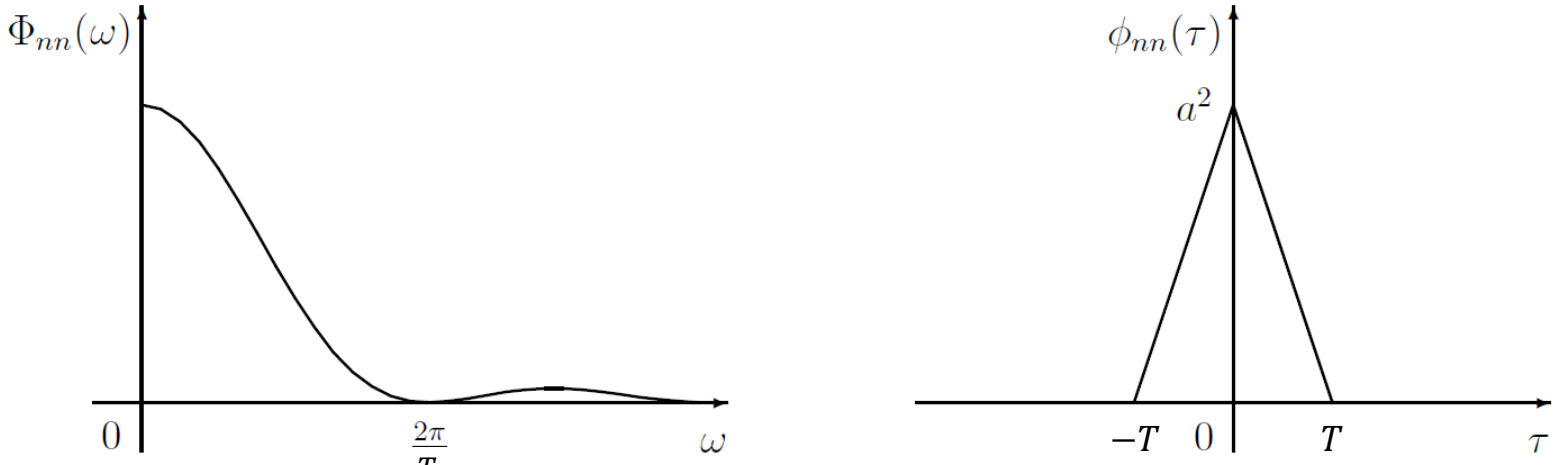
- Autocorrelation function of DRBS can be obtained by the following consideration:
  - For zero shift ($\tau = 0$) only positive products a^2 are obtained (mean value is a^2) $\rightarrow \phi_{nn}(\tau)|_{\tau=0} = a^2$
  - For small shifts ($|\tau| < T$) negative products ($-a^2$) are also obtained (their duration is proportional to the length of shift τ)
  - For $|\tau| \geq T$ positive and negative product have the same probability because shifted signal is completely independent of its unshifted version for large shifts$\rightarrow \phi_{nn}(\tau)|_{|\tau|\geq T} = 0$
- To sum up – autocorrelation of DRBS is:

$$\phi_{nn}(\tau) = \begin{cases} a^2\left[1 - \dfrac{|\tau|}{T}\right] & |\tau| < T \\ 0 & |\tau| \geq T \end{cases}$$

Discrete Random Binary Signal (DRBS)

- PSD of DRBS can be obtained by finding Fourier transform of its autocorrelation function:

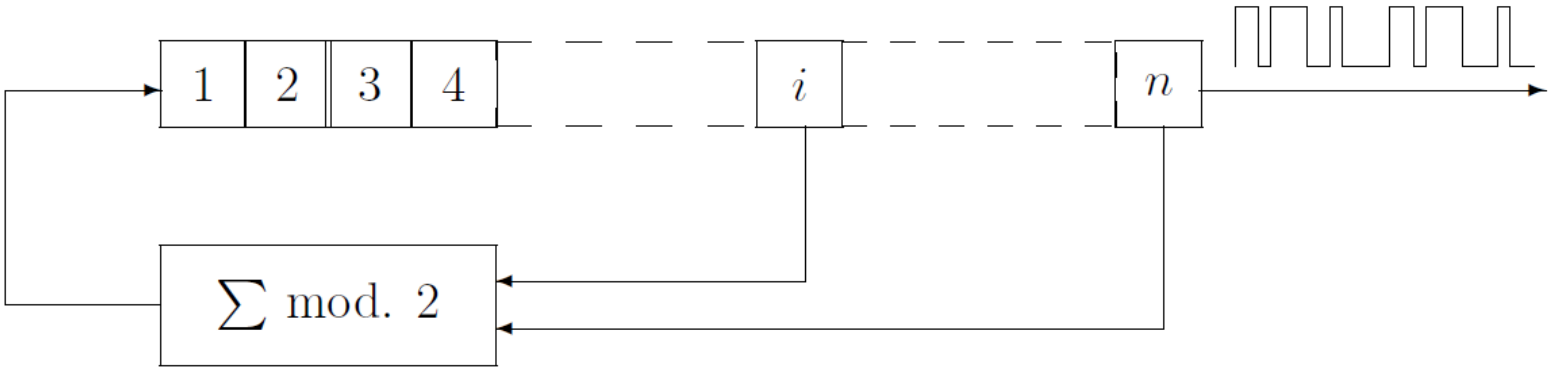
$$\Phi_{nn}(\omega) = a^2 T \left(\frac{\sin \frac{\omega T}{2}}{\frac{\omega T}{2}} \right)^2$$



- DRBS could be implemented by the shift register of infinite length with a feedback (in practice of finite length → PRBS)

Pseudo-random binary signal (PRBS)

- A finite length register is used for implementation of PRBS:
  - A shift register with n stages (storing binary information 0 or 1)
  - The contents is shifted among stages as the clock input is activated (this happens in multiples of a cycle time T)
  - The input to the first stage is generated by feeding back some stages and applying XOR operation on them
  - A periodic sequence with a period $N > n$ is obtained at the output



Pseudo-random binary signal (PRBS)

- At each cycle time a new bit (0 or 1) is obtained at the output of the register:
  - First n bits stored in the register in the beginning (initial state)
  - Then a sequence generated by feedback
  - In case of proper feedback generation (see table) all the variations of zeros and ones are obtained in the register
  - EXCEPT a forbidden one (all zeros)
  - After that the sequence is repeated
- The period N of PRBS is therefore $2^n - 1$

Number of stages (n)	Stages for feedback (inputs to XOR)	Period length (N)
2	1 and 2	3
3	1 and 3 or 2 and 3	7
4	3 and 4 or 1 and 4	15
5	2 and 5 or 3 and 5	31
6	5 and 6	63
7	4 and 7	127
8	full period not obtained (with 1 XOR)	-
9	5 and 9	511
10	7 and 10	1023
11	10 and 11	2047

Pseudo-random binary signal (PRBS)

- Any stage can be used for signal generation:
  - PRBS value is $+a$ if the stage is 1
  - PRBS value is $-a$ if the stage is 0
- Properties of a period of PRBS:
  - Each period of PRBS includes:
    - $\frac{1}{2} \cdot \frac{N+1}{2}$ impulses of length T (one half of amplitude $+a$, the other $-a$)
    - $\frac{1}{4} \cdot \frac{N+1}{2}$ impulses of length $2T$ (one half of amplitude $+a$, the other $-a$)
    - $\frac{1}{8} \cdot \frac{N+1}{2}$ impulses of length $3T$ (one half of amplitude $+a$, the other $-a$)

      ...
    - 1 negative impulse of length $(n-1)T$
    - 1 positive impulse of length nT
    - In total: $\frac{1}{2}(N+1)$ of signal values $+a$ and $\frac{1}{2}(N-1)$ of values $-a$
    - Its mean value is therefore: $E\{u(k)\} = \bar{u}(k) = \frac{a}{N}$

Pseudo-random binary signal (PRBS)

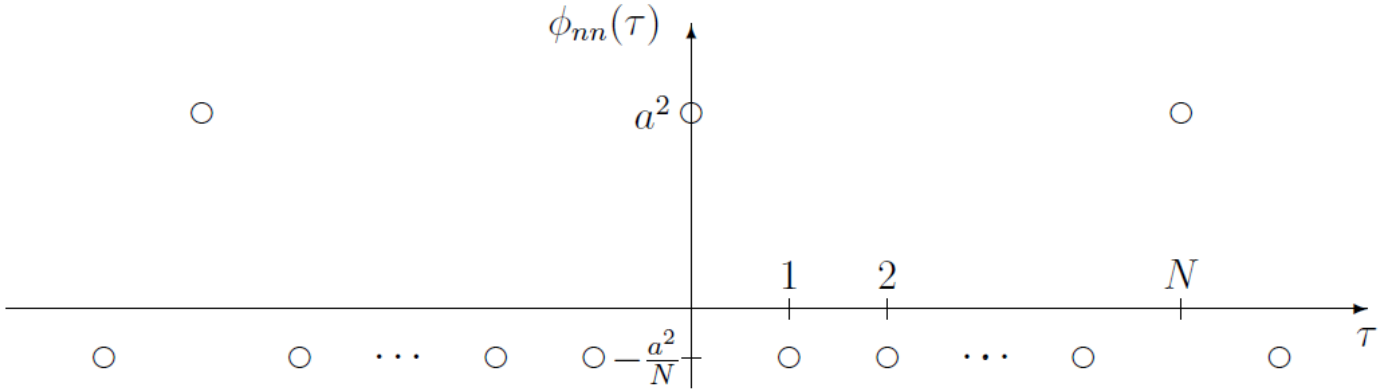
- Autocorrelation function of PRBS:
  - First autocorrelation function of discrete PRBS will be calculated:
    - Here τ_i will denote a discrete shift (the actual shift is $\tau_i T$)
    - Mean square value of the signal (for $\phi_{nn}^d(\tau_i)|_{\tau_i=0}$) is a^2
  - Due to periodicity, the same holds for $\tau_i = N, 2N, 3N, \ldots$
  - For other shifts $0 < \tau_i < N$:
    - $\frac{1}{2}(N+1)$ of negative products $-a^2$ and $\frac{1}{2}(N-1)$ of positive ones
    - $\rightarrow$ mean value of all the products is $-\frac{a^2}{N}$
  - In summary and taking into account periodicity with period N:

$$\phi_{nn}^d(\tau_i) = \begin{cases} a^2 & \tau_i = 0 \\ -\frac{a^2}{N} & 0 < \tau_i < N \end{cases}$$

$$\phi_{nn}^d(\tau_i \pm iN) = \phi_{nn}^d(\tau_i) \qquad i = 0,1,2, \ldots$$

Pseudo-random binary signal (PRBS)

- Autocorrelation function of PRBS:



- For large N autocorrelation function of PRBS tends to:

$$\phi_{nn}^d(\tau_i) = a^2\delta(\tau_i) \qquad -N < \tau_i < N$$

- Discrete PSD of PRBS:

$$\Phi_{nn}^d(m) = \sum_{\tau_i=0}^{N-1} \phi_{nn}(\tau_i) e^{-j\frac{2\pi}{N}m\tau_i} = a^2 - \frac{a^2}{N}\sum_{\tau_i=1}^{N-1} e^{-j\frac{2\pi}{N}m\tau_i}$$

$$= \begin{cases} \frac{a^2}{N} & m = 0 \\ a^2\left(1 + \frac{1}{N}\right) & 0 < m < N \end{cases}$$

The concepts of bias and consistency

- The bounds in many integrals and summations are infinite
- In practice finite bounds are used instead (the estimate of the true value is used and denoted with a hat ˆ)
- Two questions arise if stochastic disturbances are present:
  1. Is the mathematical expectation of the estimate the same as the true value? If so, the estimate is **bias-free**. The estimate is **consistent** if it improves with time (larger observation interval) and it converges to the true value as the interval of observation becomes infinite.
  2. Does the variance of the estimation error converge to 0 as the observation time converges to infinity? If so and the estimate is consistent the estimate is **consistent in the mean square (sense)**. This means that with growing observation time the accuracy and the precision of the estimates become excellent (all the estimates are in the vicinity of the true value).

The concepts of bias and consistency

- Example – estimation of the mean value:
  - Mean value of a stochastic variable x is defined as:

$$\bar{x} = E\{x(k)\} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x(k)$$

  - It is estimated using the formula:

$$\hat{\bar{x}} = \frac{1}{N} \sum_{k=1}^{N} x(k)$$

  - Is the estimate $\hat{\bar{x}}$ of $\bar{x}$ bias-free? Is it consistent? Is it consistent in the mean square?
  - Mathematical expectation of the mean-value estimate $\hat{\bar{x}}$ is:

$$E\{\hat{\bar{x}}\} = E\left\{\frac{1}{N} \sum_{k=1}^{N} x(k)\right\} = \frac{1}{N} \sum_{k=1}^{N} E\{x(k)\} = \frac{1}{N} N\bar{x} = \bar{x}$$

  - Mathematical expectation of the estimate $\hat{\bar{x}}$ is the same as the true value $\bar{x}$ even in the case of a finite observation interval N
  - The estimate is therefore **bias-free** and consequently **consistent**

The concepts of bias and consistency

- Example – estimation of the mean value:
  - The variance of the estimation error (difference between $\hat{\bar{x}}$ and $\bar{x}$):

$$E\{(\hat{\bar{x}} - \bar{x})^2\} = E\left\{\left[\frac{1}{N}\sum_{k=1}^{N}\left[x(k) - \bar{x}\right]\right]^2\right\}$$

  - If the individual observations $x(k)$ of the random variable are mutually statistically independent (white noise), the variance expression can be simplified

$$E\{(\hat{\bar{x}} - \bar{x})^2\} = E\left\{\frac{1}{N^2}\sum_{k=1}^{N}(x(k) - \bar{x})^2\right\} = \frac{1}{N^2}\sum_{k=1}^{N}E\{(x(k) - \bar{x})^2\} =$$

$$= \frac{1}{N^2}\sum_{k=1}^{N}\sigma_x^2 = \frac{1}{N^2}N\sigma_x^2 = \frac{1}{N}\sigma_x^2$$

  - The variance of the estimation error decreases with increasing number of measurements and converges to 0 in the limit case
  - The estimate is therefore **consistent in the mean square**

The concepts of bias and consistency

- Example – estimation of the variance:
  - Variance of a stochastic variable x is calculated as

$$\hat{\sigma}_x^2 = \frac{1}{N}\sum_{k=1}^{N}[x(k) - \hat{\bar{x}}]^2$$

  - Is $\hat{\sigma}_x^2$ bias-free estimate of the true variance? Is it consistent?
  - Mathematical expectation of the variance estimation:

$$E\{\hat{\sigma}_x^2\} = E\left\{\frac{1}{N}\sum_{k=1}^{N}(x(k) - \hat{\bar{x}})^2\right\}$$

  - The following holds:

$$\sum_{k=1}^{N}(x(k) - \hat{\bar{x}})^2 = \sum_{k=1}^{N}[(x(k) - \bar{x}) - (\hat{\bar{x}} - \bar{x})]^2 =$$
$$\sum_{k=1}^{N}[x(k) - \bar{x}]^2 + \sum_{k=1}^{N}(\hat{\bar{x}} - \bar{x})^2 - 2(\hat{\bar{x}} - \bar{x})\sum_{k=1}^{N}[x(k) - \bar{x}] =$$
$$\sum_{k=1}^{N}[x(k) - \bar{x}]^2 + N(\hat{\bar{x}} - \bar{x})^2 - 2(\hat{\bar{x}} - \bar{x})N(\hat{\bar{x}} - \bar{x}) =$$
$$\sum_{k=1}^{N}[x(k) - \bar{x}]^2 - N(\hat{\bar{x}} - \bar{x})^2$$

  - And: $E\{\hat{\sigma}_x^2\} = \frac{1}{N}\left[\sum_{k=1}^{N}E\{[x(k) - \bar{x}]^2\} - NE\{(\hat{\bar{x}} - \bar{x})^2\}\right] = \sigma_x^2 - \frac{1}{N}\sigma_x^2$
  - The estimate is **biased** (for finite N), but **consistent**

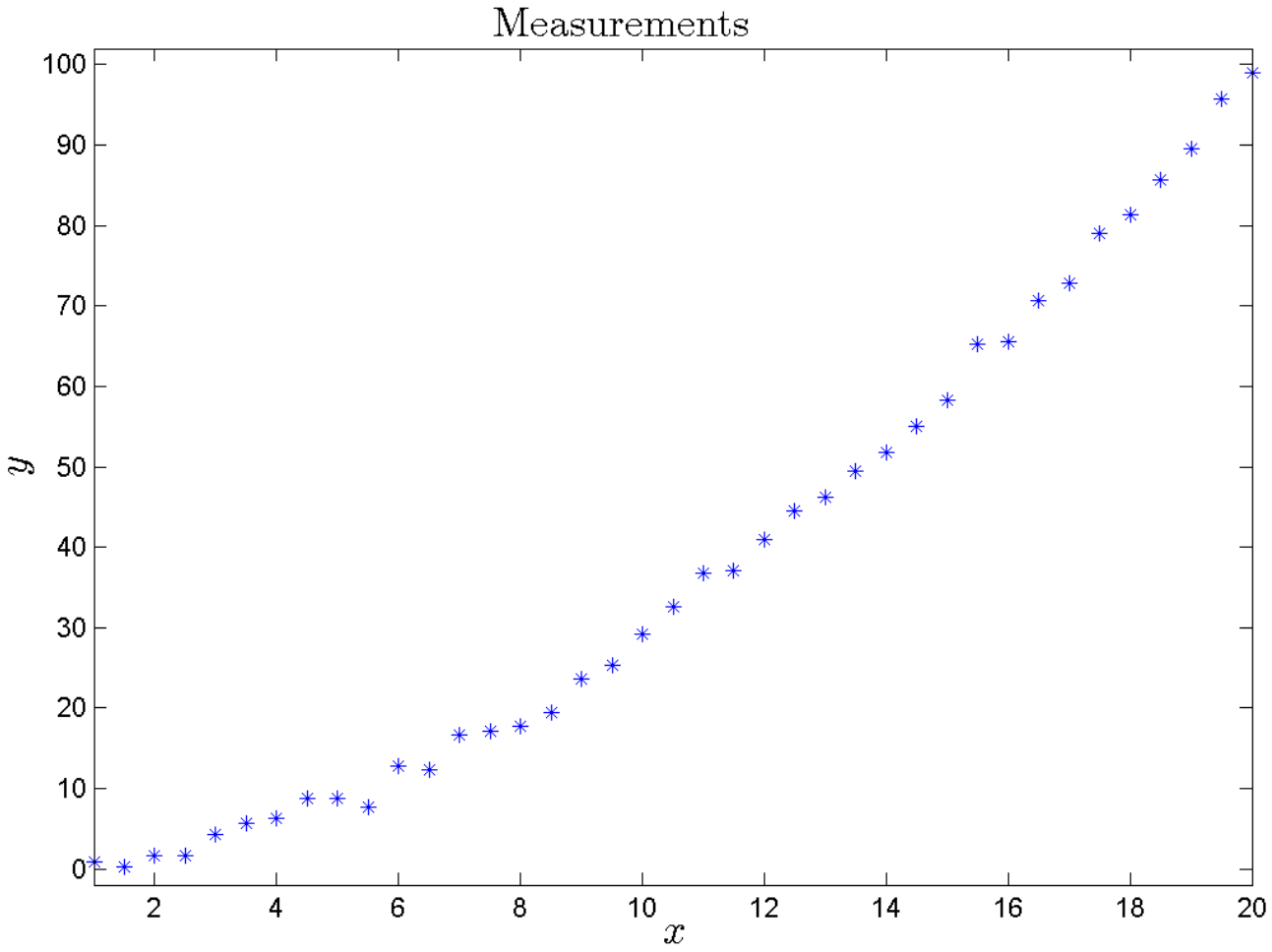
Parametric identification

- In general methods for unknown parameters determination – a very broad palette of approaches due to a large set of possible parametric models
- Parametric model:
  - A mathematical relation between measured signals and parameters
  - A very simple example: Ohm‘s law $I = \frac{U}{R}$
- Structural identification:
  - System structure determination
  - How the underlying equation (e.g. Ohm‘s law) is obtained?
- Parametric identification methods:
  - Methods for unknown parameters determination
- These two steps often performed in cycles

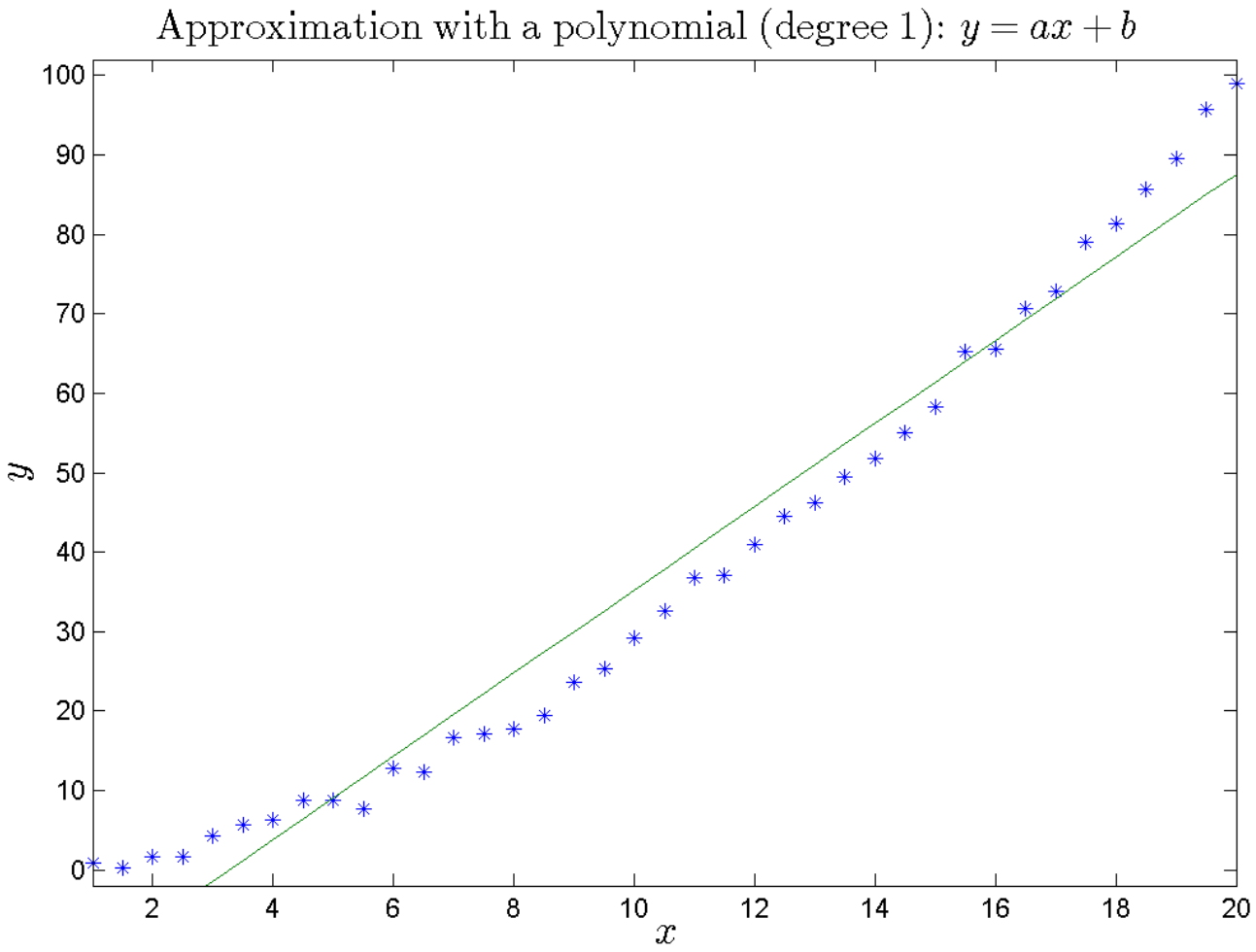
Parametric identification methods

- The problem is often rewritten as an overdetermined system of equations:
  - **More equations than unknowns**
  - All equations cannot be fulfilled simultaneously
  - If the error function is defined as a sum of squares of individual errors, the corresponding estimate is the **least-square** one (LS)
- Quadratic cost function → analytical solution
- This is a very old method:
  - First used at the end of the 18th century
  - Carl Friedrich Gauss (1777-1855)

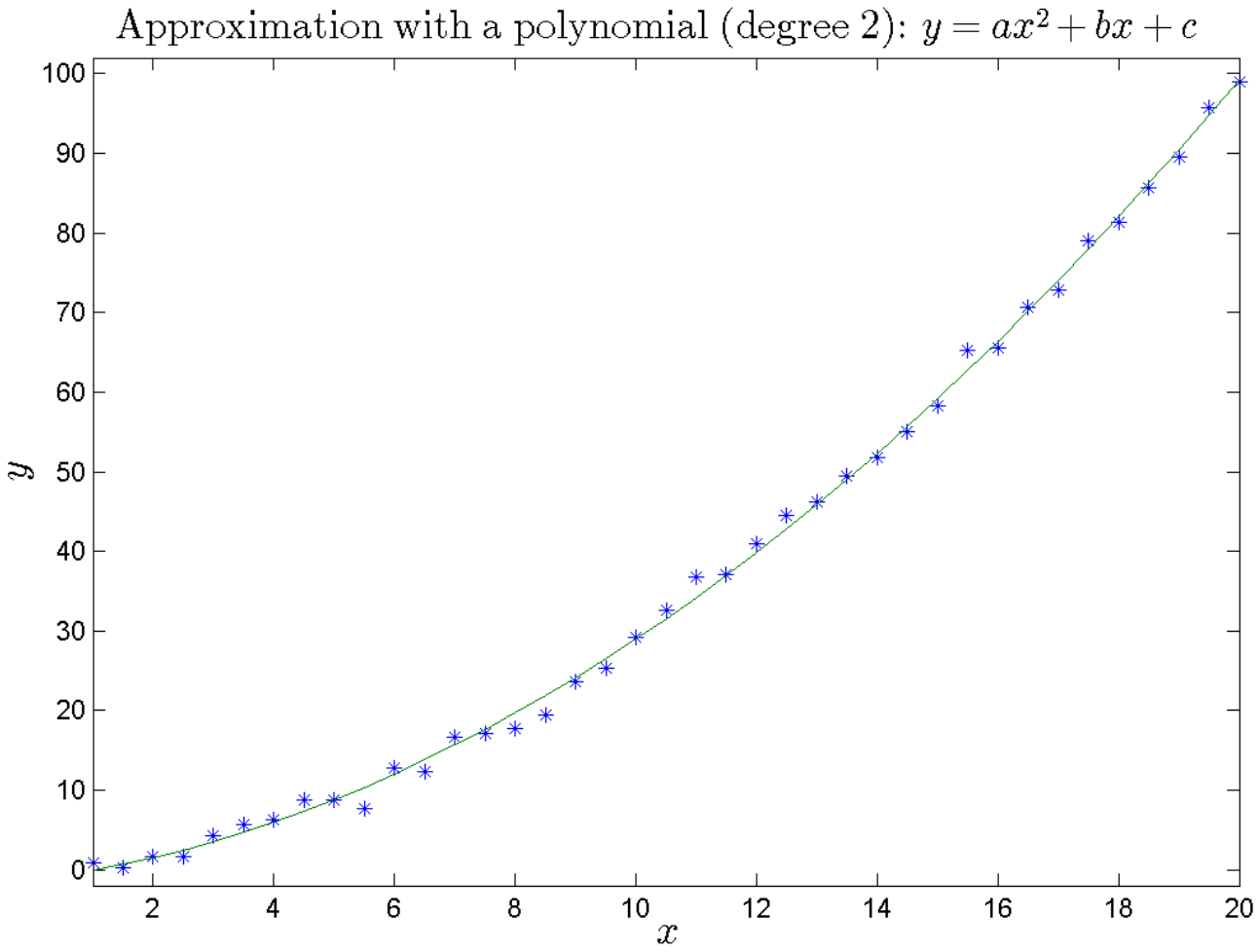
Example 1


Measurements
100
90
80
70
60
50
40
30
20
10
0
y
2
4
6
8
10
12
14
16
18
20
x

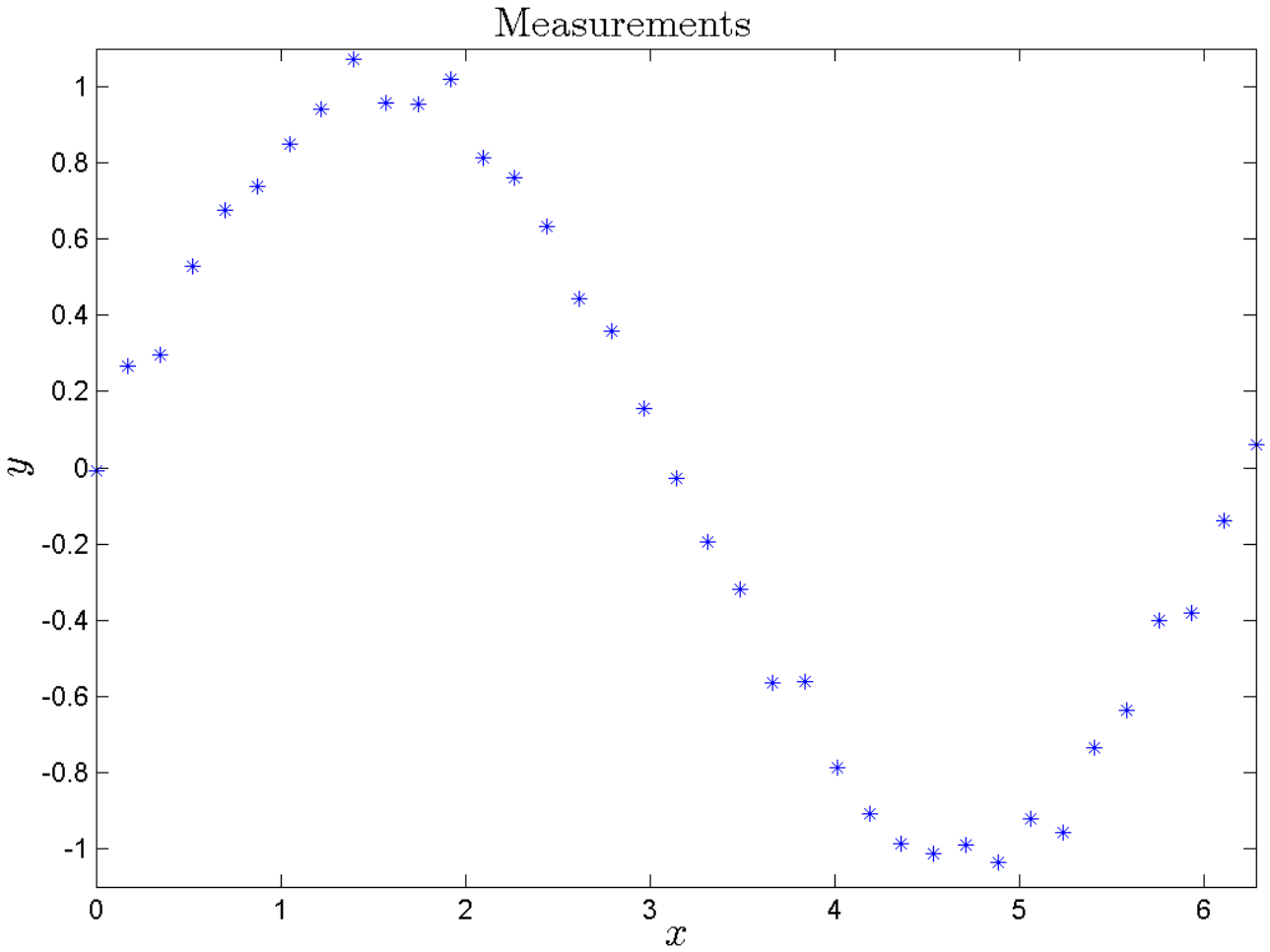
Example 1


Approximation with a polynomial (degree 1): $y = ax + b$
100
90
80
70
60
50
40
30
20
10
0
y
2
4
6
8
10
12
14
16
18
20
x

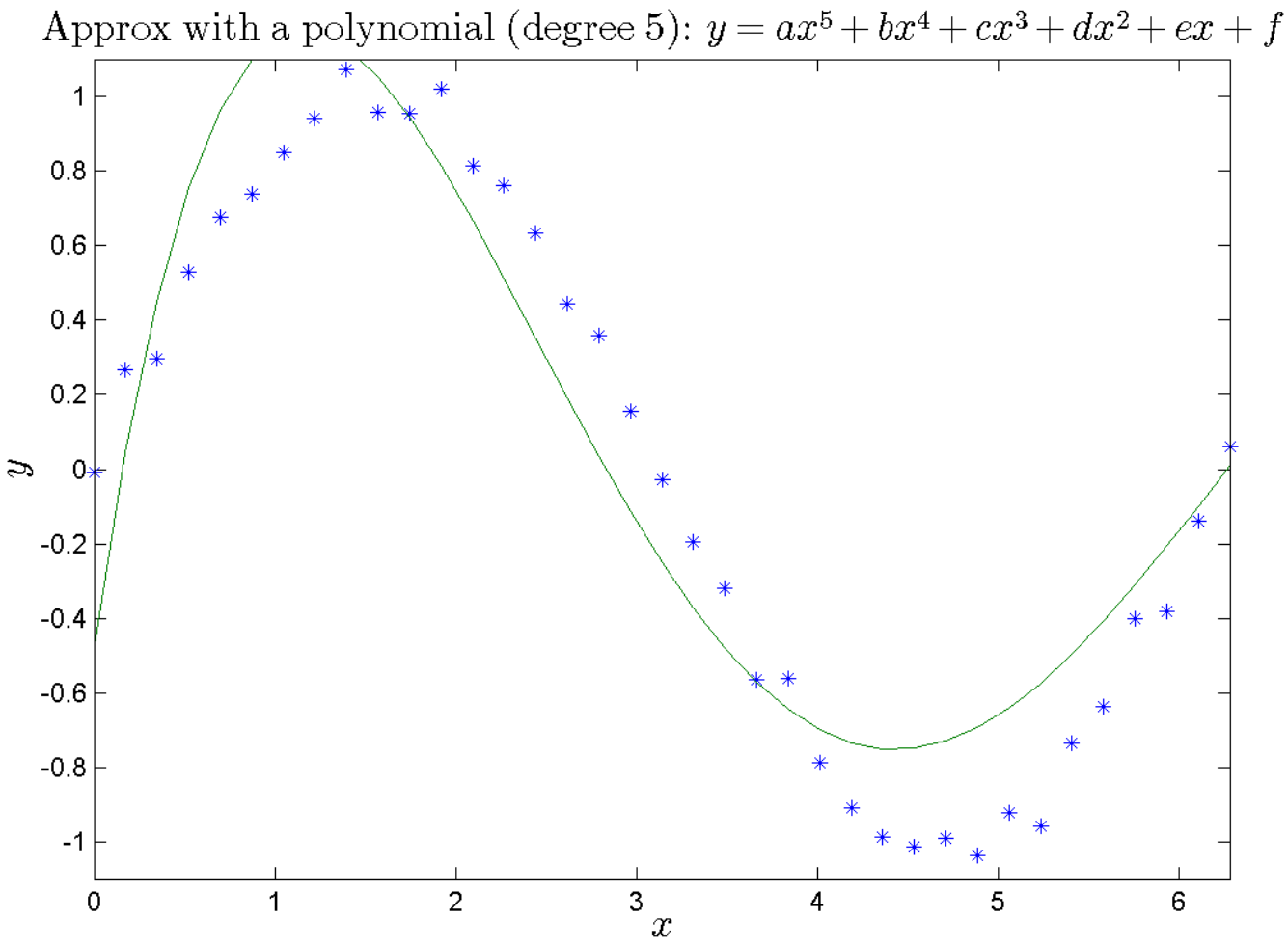
Example 1


Approximation with a polynomial (degree 2): $y = ax^2 + bx + c$
100
90
80
70
60
50
40
30
20
10
0
y
2
4
6
8
10
12
14
16
18
20
x

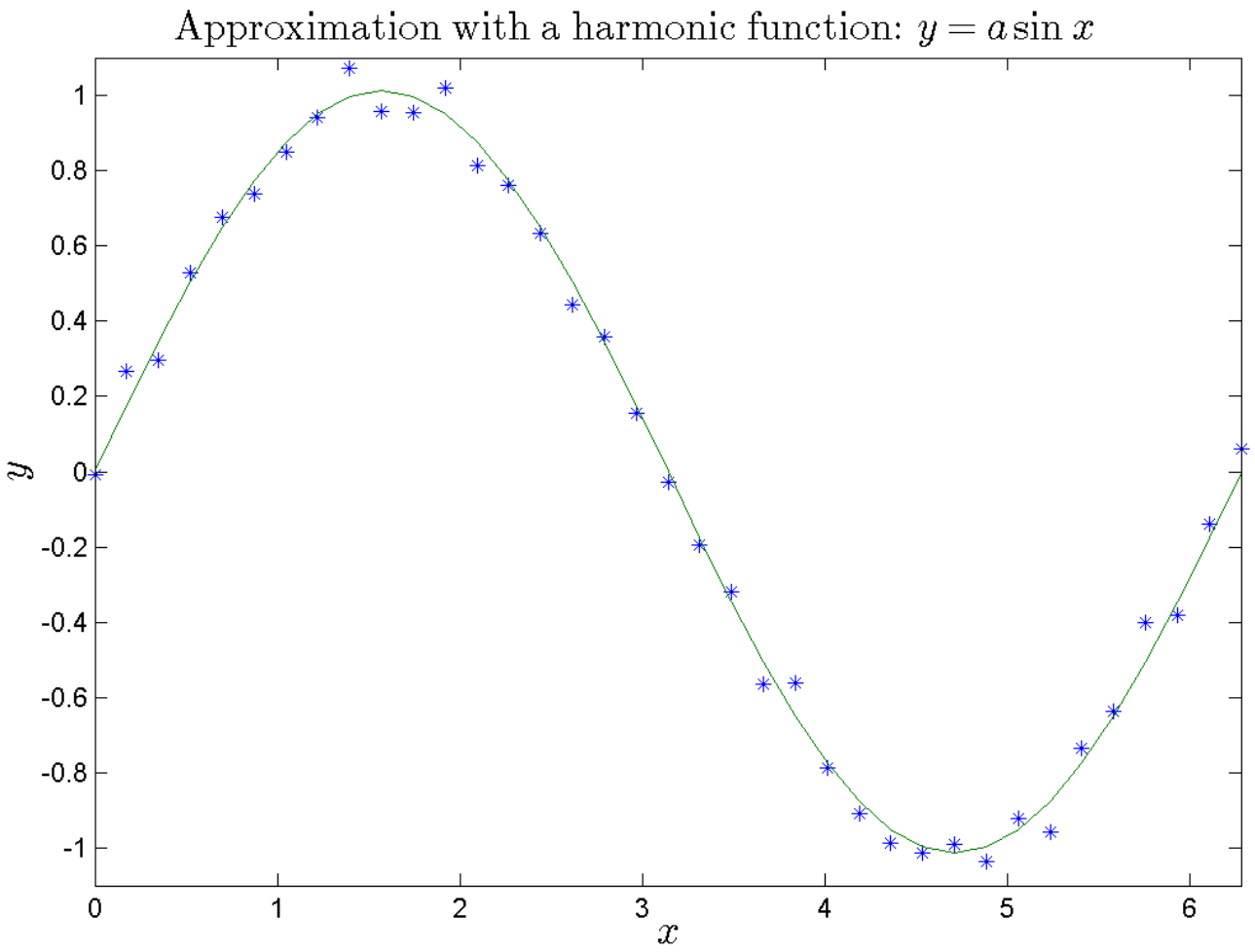
Example 2



Example 2


Approx with a polynomial (degree 5): $y = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
y
0
1
2
3
4
5
6
x

Example 2


Approximation with a harmonic function: $y = a \sin x$
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
y
0
1
2
3
4
5
6
x

Some conclusions to these two examples

- Usually the approximation improves by increasing the number of parameters, but this is not guaranteed
- Example 2 showed that a priori knowledge about the system structure ($\sin x$ in our case) often results in better approximation and lower number of parameters (only amplitude in our case)
- When referring to parametric identification sometimes the complete identification (structural and parametric) is in our mind and sometimes only the parameter estimation method:
  - structural identification – system structure determination
  - parametric identification – unknown parameters estimation, e.g. using least squares parameter estimation
  - these two steps often performed in cycles

Scalar problem

- Model of the process:
$$y_0(k) = Ku(k)$$
- Process with disturbances (disturbed measurement):
$$y(k) = y_0(k) + n(k)$$
- Process model used in estimation algorithm:
$$\hat{y}(k) = \hat{K}u(k)$$
- Only one measurement – a pair (u, y) – is enough for estimation of unknown K, but because of disturbances more measurements are used – they are indexed with argument k (k is an arbitrary independent discrete variables – very often discretised time)
- Observation error or simply error:
$$e(k) = y(k) - \hat{y}(k)$$

Scalar problem

- Errors of all N observations:

$$y(0) - \hat{K}u(0) = e(0)$$
$$y(1) - \hat{K}u(1) = e(1)$$
$$\vdots$$
$$y(N-1) - \hat{K}u(N-1) = e(N-1)$$

- The cost function is the sum of the squared errors:

$$V = \sum_{k=0}^{N-1} e^2(k) = \sum_{k=0}^{N-1} \left[y(k) - \hat{K}u(k)\right]^2$$

- The cost function has to be minimized for the parameter $\hat{K}$:

$$\frac{\partial V}{\partial \hat{K}} = -2 \sum_{k=0}^{N-1} \left[y(k) - \hat{K}u(k)\right]u(k) = 0$$

$$\hat{K} \sum_{k=0}^{N-1} u^2(k) = \sum_{k=0}^{N-1} u(k)y(k)$$

Scalar problem

- Solution:

$$\widehat{K} = \frac{\sum_{k=0}^{N-1} y(k)u(k)}{\sum_{k=0}^{N-1} u^2(k)} = \frac{\hat{\phi}_{uy}(0)}{\hat{\phi}_{uu}(0)}$$

- The second derivative is positive (!) $\rightarrow$ minimum of V:

$$\frac{\partial^2 V}{\partial \widehat{K}^2} = 2\sum_{k=0}^{N-1} u^2(k) \geq 0$$

- The solution exists if the denominator of $\widehat{K}$ not equal to 0:

$$\sum_{k=0}^{N-1} u^2(k) \neq 0$$

or

$$\hat{\phi}_{uu}(0) \neq 0$$

Scalar problem – bias? (disturbance on y)

- Mathematical expectation of the estimate:

$$E\{\widehat{K}\} = E\left\{\frac{\sum_{k=0}^{N-1}\ [y_0(k)+n(k)]u(k)}{\sum_{k=0}^{N-1}\ u^2(k)}\right\} =$$

$$= \frac{\sum_{k=0}^{N-1}\ y_0(k)u(k)}{\sum_{k=0}^{N-1}\ u^2(k)} + E\left\{\frac{\sum_{k=0}^{N-1}\ n(k)u(k)}{\sum_{k=0}^{N-1}\ u^2(k)}\right\}$$

$$= K + \frac{\sum_{k=0}^{N-1}\ E\{n(k)u(k)\}}{\sum_{k=0}^{N-1}\ u^2(k)}$$

- The estimate is bias-free if:

$$E\{n(k)u(k)\} = 0$$

  - If $u(k)$ and $n(k)$ are not correlated (cross-covariance is 0):

$$E\{n(k)u(k)\} = E\{n(k)\}E\{u(k)\}$$

    - If **mean value of either** $u(k)$ **or** $n(k)$ **is 0** and they are **not correlated**, the estimate is **bias-free**

Scalar problem – consistent in the mean square?

- Mathematical expectation of the estimate variance:

$$\sigma_K^2 = E\left\{[\widehat{K} - K]^2\right\} = \frac{E\left\{\left[\sum_{k=0}^{N-1} n(k)u(k)\right]^2\right\}}{\left[\sum_{k=0}^{N-1} u^2(k)\right]^2}$$

- Analysis of the numerator of the above expression:

$$E\left\{\sum_{k=0}^{N-1} n(k)u(k) \times \sum_{k\prime=0}^{N-1} n(k')u(k')\right\}$$

$$= \sum_{k=0}^{N-1}\sum_{k\prime=0}^{N-1} E\{n(k)n(k')u(k)u(k')\}$$

$$= \sum_{k=0}^{N-1}\sum_{k\prime=0}^{N-1} \phi_{nn}(k-k')\phi_{uu}(k-k')$$

  - k is discrete time and not only the measurement index

Scalar problem – consistent in the mean square?

- Special case – disturbance is white noise:

$$\phi_{nn}(\tau) = \Phi_{n0}\delta(\tau) = \sigma_n^2\delta(\tau)$$

  - The numerator of the variance of estimated parameter becomes:

$$\sum_{k=0}^{N-1}\sum_{k\prime=0}^{N-1}\phi_{nn}(k-k')\phi_{uu}(k-k') =$$

$$= \sigma_n^2\sum_{k=1}^{N-1}\phi_{uu}(0) = \sigma_n^2 N\phi_{uu}(0) \approx \sigma_n^2\sum_{k=0}^{N-1}u^2(k)$$

equality in case of infinite N

  - The variance expression simplifies to:

$$\sigma_K^2 \approx \frac{\sigma_n^2}{\sum_{k=0}^{N-1}u^2(k)} \approx \frac{1}{N}\times\frac{\sigma_n^2}{\sigma_u^2}$$

    - If mean value of u is equal to 0, the $\approx$ signs in the above equation change to $=$

Scalar problem – consistent in the mean square?

- Special case – u is white noise:

$$\phi_{uu}(\tau) = \Phi_{u0}\delta(\tau) = \sigma_u^2\delta(\tau)$$

  - Expression in the numerator of the variance of estimated parameter becomes:

$$\sum_{k=0}^{N-1}\sum_{k\prime=0}^{N-1}\phi_{nn}(k-k')\phi_{uu}(k-k') =$$

$$= \sigma_u^2 N\phi_{nn}(0) \approx \sigma_u^2 N\sigma_n^2$$

    - If mean value of n is equal to 0, the $\approx$ sign in the above equation changes to $=$

  - Variance of the estimated parameter:

$$\sigma_K^2 \approx \frac{N\sigma_u^2\sigma_n^2}{[\sum_{k=0}^{N-1} u^2(k)]^2} \approx \frac{1}{N} \times \frac{\sigma_n^2}{\sigma_u^2}$$

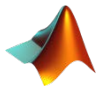
    - If mean value of u is equal to 0, the $\approx$ signs in the above equation change to $=$

Scalar problem – consistent in the mean square?

- If either u or n is a white noise and both signals are zero-mean:

$$\sigma_K^2 \approx \frac{1}{N} \times \frac{\sigma_n^2}{\sigma_u^2}$$

- Variance of the estimated parameter error is directly proportional to the variance of the disturbance and inversely proportional to the variance (power) of the excitation. Moreover, the variance tends to 0 as the number of observations increases, and therefore the estimate is **consistent in the mean square**
- This is a quite general rule in system identification:
  - The variance of the estimated parameter increases with the increasing power/energy of disturbance
  - The variance of the estimated parameter decreases with the increasing power/energy of the excitation



Vector problem

- Several unknown parameters are estimated simultaneously
- The process (a static mapping between an input u and an output y_0) can be described by:

$$y_0 = a_1 f_1(u) + a_2 f_2(u) + \cdots + a_n f_n(u)$$

- Very important fact: y_0 **depends linearly** on a_i, although $f_i(u)$ can be arbitrary known functions (also nonlinear)
- The above n parameters can be estimated if n measurements are available
- If more measurements exist, the estimation system is overdetermined – least squares (LS) parameter estimation is used – the sum of squared errors between the measurements and the model outputs is minimised
- Special case of the vector problem – regression polynomials estimation

Vector problem

- Mathematical model of a regression polynomial of order q:

$$y_0 = K_0 + K_1 u + K_2 u^2 + \cdots + K_q u^q$$

- Again N measurements are available:

$$y_0(k) = K_0 + K_1 u(k) + K_2 u^2(k) + \cdots + K_q u^q(k),$$
$$k = 0,1,\cdots,N-1$$

- Again the measurements are corrupted by noise:

$$y(k) = y_0(k) + n(k)$$

- The model of the process used in estimation algorithm:

$$\hat{y}(k) = \hat{K}_0 + \hat{K}_1 u(k) + \hat{K}_2 u^2(k) + \cdots + \hat{K}_q u^q(k)$$

- The error is again the output error:

$$e(k) = \hat{y}(k) - y(k)$$

Vector problem

- Process observed at different times:

$$y(0) = K_0 + K_1 u(0) + K_2 u^2(0) + \cdots + K_q u^q(0) + n(0)$$
$$y(1) = K_0 + K_1 u(1) + K_2 u^2(1) + \cdots + K_q u^q(1) + n(1)$$
$$\vdots$$
$$y(N-1) = K_0 + K_1 u(N-1) + \cdots + K_q u^q(N-1) + n(N-1)$$

- Compact description of the process equation (output $\mathbf{y}$ depends linearly on $\mathbf{K}$):

$$\mathbf{y} = \mathbf{UK} + \mathbf{n}$$

- Output error evaluated at N different times :

$$y(0) - \hat{K}_0 - \hat{K}_1 u(0) - \hat{K}_2 u^2(0) - \cdots - \hat{K}_q u^q(0) = e(0)$$
$$y(1) - \hat{K}_0 - \hat{K}_1 u(1) - \hat{K}_2 u^2(1) - \cdots - \hat{K}_q u^q(1) = e(1)$$
$$\vdots$$
$$y(N-1) - \hat{K}_0 - \hat{K}_1 u(N-1) - \cdots - \hat{K}_q u^q(N-1) = e(N-1)$$

- Compact description of the output error:

$$\mathbf{y} - \mathbf{U}\hat{\mathbf{K}} = \mathbf{e}$$

Vector problem

- The vectors and matrices used in compact descriptions:

$$\mathbf{U} = \begin{bmatrix} 1 & u(0) & u^2(0) & \cdots & u^q(0) \\ 1 & u(1) & u^2(1) & \cdots & u^q(1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u(N-1) & u^2(N-1) & \cdots & u^q(N-1) \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix} \quad \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}(0) \\ \hat{y}(1) \\ \vdots \\ \hat{y}(N-1) \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e(0) \\ e(1) \\ \vdots \\ e(N-1) \end{bmatrix}$$

$$\mathbf{n} = \begin{bmatrix} n(0) \\ n(1) \\ \vdots \\ n(N-1) \end{bmatrix} \quad \mathbf{K} = \begin{bmatrix} K_0 \\ K_1 \\ \vdots \\ K_q \end{bmatrix} \quad \hat{\mathbf{K}} = \begin{bmatrix} \hat{K}_0 \\ \hat{K}_1 \\ \vdots \\ \hat{K}_q \end{bmatrix}$$

Vector problem

- Cost function – sum of squared errors in vector-matrix form:

$$V = \mathbf{e}^T\mathbf{e} = \left[\mathbf{y} - \mathbf{U}\hat{\mathbf{K}}\right]^T\left[\mathbf{y} - \mathbf{U}\hat{\mathbf{K}}\right] =$$
$$= \left[\mathbf{y}^T - \hat{\mathbf{K}}^T\mathbf{U}^T\right]\left[\mathbf{y} - \mathbf{U}\hat{\mathbf{K}}\right] =$$
$$= \mathbf{y}^T\mathbf{y} - \hat{\mathbf{K}}^T\mathbf{U}^T\mathbf{y} - (\mathbf{U}^T\mathbf{y})^T\hat{\mathbf{K}} + \hat{\mathbf{K}}^T\mathbf{U}^T\mathbf{U}\hat{\mathbf{K}}$$

- Taking the first derivative of the cost function with respect to the parameter vector $\hat{\mathbf{K}}$ yields (and equating it to zero):

$$\frac{\partial V}{\partial \hat{\mathbf{K}}} = -2\mathbf{U}^T\mathbf{y} + 2\mathbf{U}^T\mathbf{U}\hat{\mathbf{K}} = -2[\mathbf{U}^T\mathbf{y} - \mathbf{U}^T\mathbf{U}\hat{\mathbf{K}}] = \mathbf{0}$$

- The solution:

$$\hat{\mathbf{K}} = [\mathbf{U}^T\mathbf{U}]^{-1}\mathbf{U}^T\mathbf{y}$$

- The solution is a minimum of V due to:

$$\frac{\partial^2 V}{\partial \hat{\mathbf{K}}^2} = 2\mathbf{U}^T\mathbf{U} \quad \rightarrow \text{a } \mathbf{positive\ definite\ matrix}$$

Vector problem – bias? (disturbance on y)

- The influence of disturbance $\mathbf{n}$ to the estimate $\widehat{\mathbf{K}}$:

$$\widehat{\mathbf{K}} = [\mathbf{U}^T\mathbf{U}]^{-1}\mathbf{U}^T(\mathbf{U}\mathbf{K} + \mathbf{n}) = \mathbf{K} + [\mathbf{U}^T\mathbf{U}]^{-1}\mathbf{U}^T\mathbf{n}$$

- Mathematical expectation of the estimate:

$$E\{\widehat{\mathbf{K}}\} = \mathbf{K} + E\{[\mathbf{U}^T\mathbf{U}]^{-1}\mathbf{U}^T\mathbf{n}\}$$

- If the input $\mathbf{U}$ and the noise $\mathbf{n}$ are not correlated, mathematical expectation of the above product can be rewritten as a product of mathematical expectations (mathematical expectation operator of a deterministic term $[\mathbf{U}^T\mathbf{U}]^{-1}\mathbf{U}^T$ omitted)

$$E\{\widehat{\mathbf{K}}\} = \mathbf{K} + [\mathbf{U}^T\mathbf{U}]^{-1}\mathbf{U}^T E\{\mathbf{n}\}$$

- The estimate is therefore **bias-free** if the input is not correlated with a disturbance and the latter is zero-mean

Vector problem – consistent in the mean square?

- Covariance matrix of the estimation error:

$$\mathrm{cov}[\widehat{\mathbf{K}} - \mathbf{K}] = E\left\{[\widehat{\mathbf{K}} - \mathbf{K}] \times [\widehat{\mathbf{K}} - \mathbf{K}]^T\right\}$$
$$= E\{[\mathbf{U}^T\mathbf{U}]^{-1}\mathbf{U}^T\mathbf{n}\mathbf{n}^T\mathbf{U}[\mathbf{U}^T\mathbf{U}]^{-1}\}$$

- If the input $\mathbf{U}$ and the noise $\mathbf{n}$ are not correlated:

$$\mathrm{cov}[\widehat{\mathbf{K}} - \mathbf{K}] = [\mathbf{U}^T\mathbf{U}]^{-1}\mathbf{U}^T E\{\mathbf{n}\mathbf{n}^T\}\mathbf{U}[\mathbf{U}^T\mathbf{U}]^{-1}$$

- Mathematical expectation of the product of shifted samples of the disturbance $\rightarrow$ auto-correlation of the disturbance:

$$E\{\mathbf{n}\mathbf{n}^T\} = \begin{bmatrix} \phi_{nn}(0) & \phi_{nn}(1) & \cdots & \phi_{nn}(N-1) \\ \phi_{nn}(1) & \phi_{nn}(0) & \cdots & \phi_{nn}(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{nn}(N-1) & \phi_{nn}(N-2) & \cdots & \phi_{nn}(0) \end{bmatrix}$$

Vector problem – consistent in the mean square?

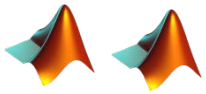
- If disturbance is white noise, its auto-correlation is a delta-impulse:

$$\phi_{nn}(\tau) = \Phi_{n0}\delta(\tau)$$
$$E\{\mathbf{n}\mathbf{n}^T\} = \Phi_{n0}\mathbf{I}$$

- Then:

$$\mathrm{cov}[\widehat{\mathbf{K}} - \mathbf{K}] = \Phi_{n0}[\mathbf{U}^T\mathbf{U}]^{-1} = \frac{1}{N}\Phi_{n0}\left[\frac{1}{N}\mathbf{U}^T\mathbf{U}\right]^{-1}$$

- The expression in the brackets is close to input signal (including its powers) covariance matrix $\rightarrow$ finite matrix, $\Phi_{n0} \rightarrow$ finite matrix
- The covariance of the estimation error converges to 0 if the number of observations N converges to infinity – the estimate is consistent in the mean square



Dilemmas about (non)linearity of the parameter estimation problem

- Parameter estimation problem of an overdetermined system can be always presented in the following vector form:

$$\mathbf{f}(\mathbf{d}, \boldsymbol{\Theta}) = 0$$

  - Vector $\mathbf{d}$ includes **known** data, e.g. measurements
  - Vector $\boldsymbol{\Theta}$ includes **unknown** data, e.g. model parameters
  - Nonlinear mapping $\mathbf{f}$ gives the relation between vectors $\mathbf{d}$ and $\boldsymbol{\Theta}$
- Very often, the data $\mathbf{d}$ can be split into two parts to obtain:

$$\mathbf{y} = \mathbf{g}(\mathbf{x}, \boldsymbol{\Theta})$$

  - The part of known data, collected in $\mathbf{y}$, is called output of the model – in general not related to the „classical“ output of the process
  - The part of known data, collected in $\mathbf{x}$, is called input of the model
- Parameter estimation problem is said to be **linear in parameters** if it can be reformulated as:

$$\mathbf{y} = \mathbf{H}(\mathbf{x})\boldsymbol{\Theta}$$

Dilemmas about (non)linearity of the parameter estimation problem

- In the previous equation $\mathbf{y}$ is a known vector, $\boldsymbol{\Theta}$ an unknown vector, and $\mathbf{H}(\mathbf{x})$ a known matrix of appropriate dimensions
- The system, described by $\mathbf{y} = \mathbf{H}(\mathbf{x})\boldsymbol{\Theta}$, is in general nonlinear (due to nonlinear relation $\mathbf{H}$ over the vector $\mathbf{x}$), but it is linear with respect to (w.r.t.) the estimated vector $\boldsymbol{\Theta}$
- Example: The system $y = kx^2$ (where k is estimated based on measured x and y) is nonlinear w.r.t. x, but linear w.r.t. k
- Example: The system $y = A\sin(\omega t + \varphi)$ (where A and φ are estimated based on measured t and y, and known frequency ω) is linear w.r.t. A and nonlinear w.r.t. φ. The system is not linear in parameters. But after the system is reformulated:

$$y = A\sin\omega t\cos\varphi + A\cos\omega t\sin\varphi = \begin{bmatrix}\sin\omega t & \cos\omega t\end{bmatrix}\begin{bmatrix}A\cos\varphi \\ A\sin\varphi\end{bmatrix}$$

  - The problem becomes linear w.r.t new parameters $A\cos\varphi$ and $A\sin\varphi$

Dilemmas about (non)linearity of the parameter estimation problem

- If the parameter estimation problem is linear in parameters, it can be presented as:

  $$known\ quantity = known\ quantity \times unknown\ parameter$$

  - As already said, the quantity on the left-hand (right-hand) side of the equation is not necessarily the process output (input)
  - Example: The system is given by:

    $$y(k) = Ku^2(k) + \sqrt{v(k)}$$

    - $u(k)$ and $v(k)$ are measured inputs, $y(k)$ is a measured output
    - Parameter estimation problem (for parameter K) is rewritten as:

      $$y(k) - \sqrt{v(k)} = u^2(k) \times K$$

    - The output of the parameter estimation model is $\left(y(k) - \sqrt{v(k)}\right)$
- When estimating parameters all the available a priori knowledge and measured data should be considered $\rightarrow$ overdetermined system – more equations (data) than unknowns (parameters)

Some conclusions

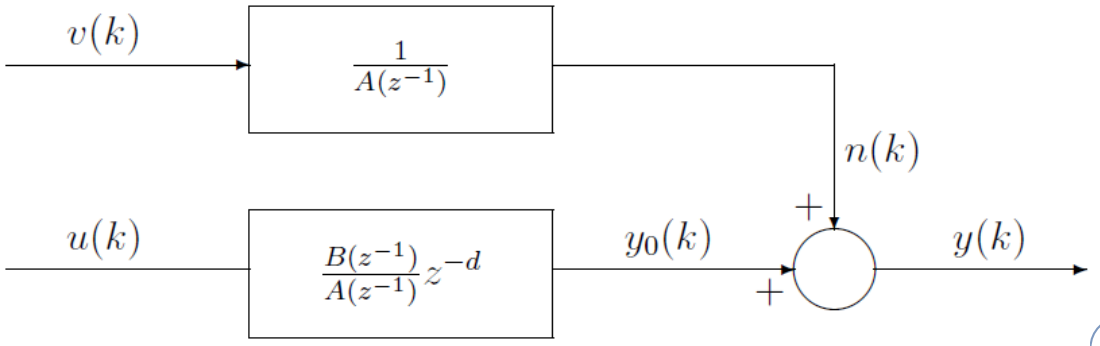
- Least squares parameter estimation is a very useful tool for solving an overdetermined system of equations, especially if unknowns appear linearly, i.e. the problem is linear in parameters
- The method is often used for estimation of unknown parameters of static systems (linear regression)
- If disturbances and independent variables (model inputs) are not correlated and the mean value of disturbances is zero, the estimate is bias-free (and therefore consistent)
- If, moreover, the disturbances are mutually statistically independent, the estimate is consistent in the mean square
- The method can be and will be extended to estimation of parameters of dynamic systems but of course some assumptions will have to be made when dealing with bias, consistency and consistency in the mean square

LS parameter estimation of discrete-time dyn. systems

- It will be assumed that the process can be described by:

$$G_p(z) = \frac{B(z^{-1})}{A(z^{-1})} z^{-d} = \frac{b_1 z^{-1} + \cdots + b_n z^{-n}}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}} z^{-d}$$

  - d – time delay – should be known *a priori*
  - n – system order – should be known *a priori*
  - a_i and b_i $(i = 1, \ldots, n)$ – system parameters estimated by least squares (LS) parameter estimation
- The output is corrupted by noise n, obtained from noise v by filtering it through a noise filter $G_n(z) = \frac{1}{A(z^{-1})}$
  - No assumption about $v(k)$ at this point



LS parameter estimation of discrete-time dyn. systems

- No special requirements posed to the properties of $v(k)$ for now
- The special form of the noise filter will be discussed later
- The z transform of the disturbed process output:

$$y(z) = \frac{B(z^{-1})}{A(z^{-1})} z^{-d} u(z) + \frac{1}{A(z^{-1})} v(z)$$

- This description can be rewritten:

$$A(z^{-1})y(z) = B(z^{-1})z^{-d}u(z) + v(z)$$

$$(1 + a_1 z^{-1} + \cdots + a_n z^{-n})y(z) = (b_1 z^{-1} + \cdots + b_n z^{-n})z^{-d}u(z) + v(z)$$

- Finally the time domain version of the equation is obtained after inverse z transform has been performed:

$$y(k) + a_1 y(k-1) + a_2 y(k-2) + \cdots + a_n y(k-n) =$$
$$= b_1 u(k-d-1) + \cdots + b_n u(k-d-n) + v(k)$$

  - Measurements known, noise unknown, parameters estimated

LS parameter estimation of discrete-time dyn. systems

- The equation in the vector form

$$y(k) = \boldsymbol{\psi}^T(k)\boldsymbol{\theta} + v(k)$$

  - $\boldsymbol{\psi}^T(k) = [-y(k-1), \dots, -y(k-n), u(k-d-1), \dots, u(k-d-n)]$
  - $\boldsymbol{\theta}^T = [a_1, \dots, a_n, b_1, \dots, b_n]$ – vector of unknown parameters
- Identification task: to estimate parameters a_i and b_i from measurements of input ($u(k)$) and output ($y(k)$) process signals
- Difference from the vector parameter estimation problem for the static case – the measurements in the regression vector $\boldsymbol{\psi}(k)$ not obtained in the same time instant
- The model used in the parameter estimation algorithm:

$$\hat{y}(k) = \boldsymbol{\psi}^T(k)\hat{\boldsymbol{\theta}} = -\hat{a}_1 y(k-1) - \cdots - \hat{a}_n y(k-n) + \\ + \hat{b}_1 u(k-d-1) + \cdots + \hat{b}_n u(k-d-n)$$

  - $\hat{\boldsymbol{\theta}}^T = [\hat{a}_1, \dots, \hat{a}_n, \hat{b}_1, \dots, \hat{b}_n]$ – parameter estimate

LS parameter estimation of discrete-time dyn. systems

- Model output $\hat{y}$ – in the literature often referred to as one-step-ahead prediction: prediction of model output in time instant k based on measurement in time instants $k-1, k-2, \ldots k-d-n$
- Z transform of the previous equation – for calculation of $\hat{y}(k)$:

$$\hat{y}(z) = [1 - \hat{A}(z^{-1})]y(z) + \hat{B}(z^{-1})z^{-d}u(z)$$

  - $\hat{A}(z^{-1}) = 1 + \hat{a}_1 z^{-1} + \cdots + \hat{a}_n z^{-n}$
  - $\hat{B}(z^{-1}) = \hat{b}_1 z^{-1} + \cdots + \hat{b}_n z^{-n}$
- Error is the difference between outputs $y(k)$ and $\hat{y}(k)$:

$$e(k) = y(k) - \hat{y}(k)$$

- The relation between the parameter estimate and the error:

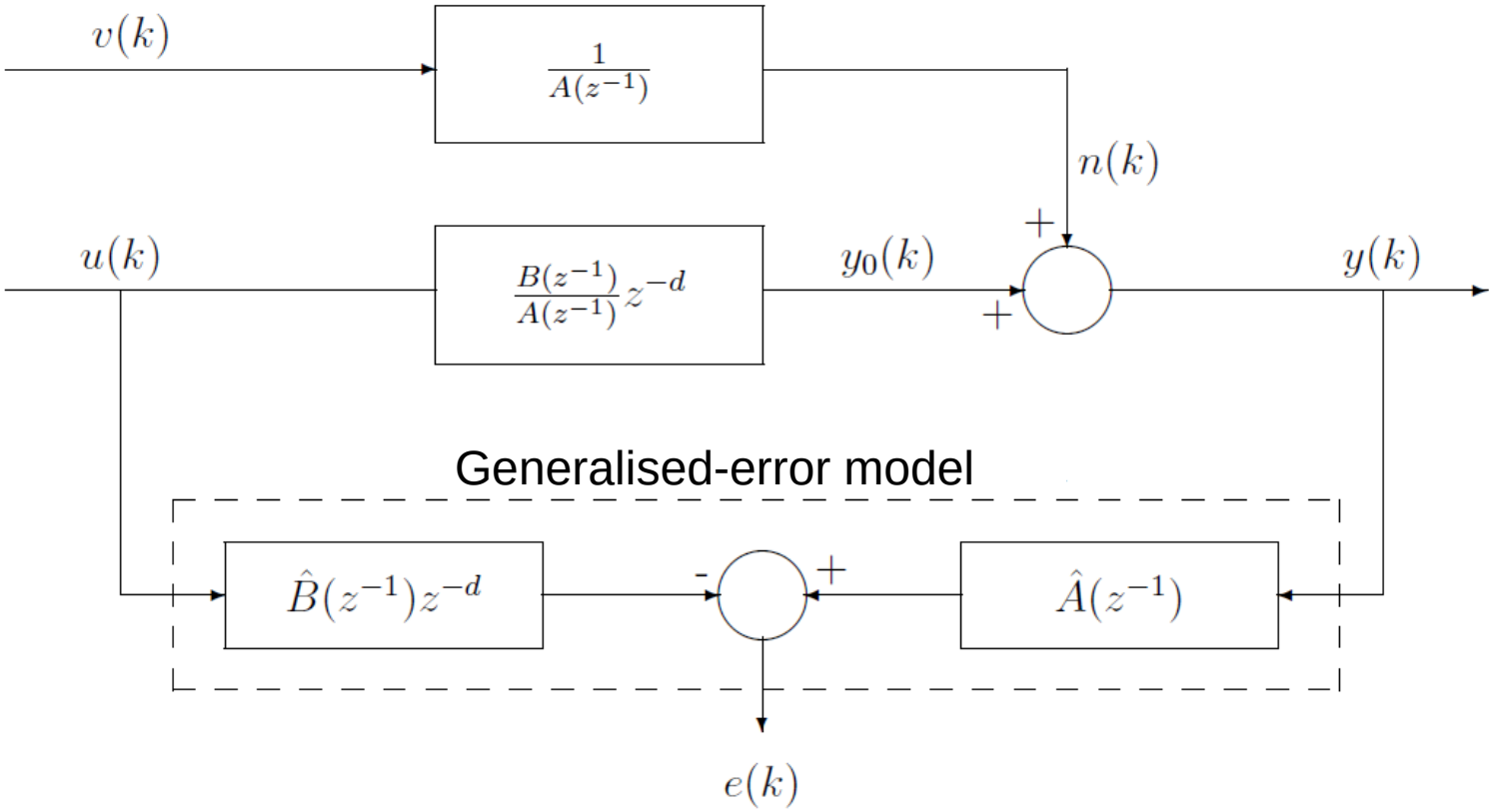
$$e(k) = y(k) - \boldsymbol{\psi}^T(k)\hat{\boldsymbol{\theta}}$$

- Taking into account $\hat{y}(z)$, the z transform of the error becomes:

$$e(z) = \hat{A}(z^{-1})y(z) - \hat{B}(z^{-1})z^{-d}u(z)$$

LS parameter estimation of discrete-time dyn. systems

- The error $e(k)$ is in fact a generalised one – obtained by „filtering“ process input and output through the model ($\hat{A}$ and $\hat{B}z^{-d}$)
- The schematic representation of the error model:



LS parameter estimation of discrete-time dyn. systems

- The system of equations for determination of parameter estimates $\hat{\boldsymbol{\theta}}$ is obtained by rewriting $e(k) = y(k) - \boldsymbol{\psi}^T(k)\hat{\boldsymbol{\theta}}$ in different observation times (N equations are thus obtained):
  - The number of equations N has to be at least as large as the number of estimated parameters ($2n$), but in practice much larger
  - The generalised error in the current time sample k and in $N-1$ past time samples:

$$
\begin{aligned}
e(k-N+1) &= y(k-N+1) - \boldsymbol{\psi}^T(k-N+1)\hat{\boldsymbol{\theta}} \\
e(k-N+2) &= y(k-N+2) - \boldsymbol{\psi}^T(k-N+2)\hat{\boldsymbol{\theta}} \\
&\vdots \\
e(k-1) &= y(k-1) - \boldsymbol{\psi}^T(k-1)\hat{\boldsymbol{\theta}} \\
e(k) &= y(k) - \boldsymbol{\psi}^T(k)\hat{\boldsymbol{\theta}}
\end{aligned}
$$

- A more compact form of the system description is obtained using vector-matrix representation

LS parameter estimation of discrete-time dyn. systems

- Vectors $\mathbf{y}$, $\mathbf{e}$, $\mathbf{v}$ and $\boldsymbol{\psi}$:

$$\mathbf{y} = \begin{bmatrix} y(k-N+1) \\ y(k-N+2) \\ \vdots \\ y(k-1) \\ y(k) \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e(\mathrm{k}-N+1) \\ e(k-N+2) \\ \vdots \\ e(k-1) \\ e(k) \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} v(k-N+1) \\ v(k-N+2) \\ \vdots \\ v(k-1) \\ v(k) \end{bmatrix}$$

$$\boldsymbol{\psi}^T(k) = [-y(k-1) \quad \dots \quad -y(k-n) \quad u(k-1-d) \quad \dots \quad u(k-n-d)]$$

- Matrix $\boldsymbol{\Psi}$:

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\psi}^T(k-N+1) \\ \boldsymbol{\psi}^T(k-N+2) \\ \vdots \\ \boldsymbol{\psi}^T(k-1) \\ \boldsymbol{\psi}^T(k) \end{bmatrix}$$

$$= \begin{bmatrix} -y(k-N) & \dots & -y(k-N-n+1) & u(k-N-d) & \dots & u(k-N-n-d+1) \\ \vdots & & \vdots & \vdots & & \vdots \\ -y(k-1) & \dots & -y(k-n) & u(k-1-d) & \dots & u(k-n-d) \end{bmatrix}$$

LS parameter estimation of discrete-time dyn. systems

- The generalised error vector given in the vector-matrix form:

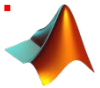
$$\mathbf{y} - \boldsymbol{\Psi}\hat{\boldsymbol{\theta}} = \mathbf{e}$$

- Similarly the process equation $y(k) - \boldsymbol{\psi}^T(k)\boldsymbol{\theta} = v(k)$ can be rewritten in the same N time instants:

$$\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\theta} = \mathbf{v}$$

- The above equations the same as in the case of the vector problem ($\mathbf{U} \to \boldsymbol{\Psi}$, $\mathbf{K} \to \boldsymbol{\theta}$, $\mathbf{n} \to \mathbf{v}$) where the solution was:

$$\hat{\boldsymbol{\theta}} = [\boldsymbol{\Psi}^T\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^T\mathbf{y}$$



- Matrix $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ is always symmetrical and square of dimensions $2n \times 2n$ (irrespective of the number of observations):
- The inverse of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ has to exist $\Leftrightarrow$ $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ of full rank $\Leftrightarrow$ $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ is non-singular $\Leftrightarrow$ its determinant different from zero:
  - This is achieved by appropriate excitation – input signal has to be „rich“ enough – in terms of its frequency content

Bias, consistency and convergence of estimates

- Introducing $\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\theta} + \mathbf{v}$ into the parameter estimation equation:
$$\hat{\boldsymbol{\theta}} = [\boldsymbol{\Psi}^T\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^T(\boldsymbol{\Psi}\boldsymbol{\theta} + \mathbf{v}) = \boldsymbol{\theta} + [\boldsymbol{\Psi}^T\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^T\mathbf{v}$$
- Mathematical expectation of the parameter estimate:
$$E\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta} + E\{[\boldsymbol{\Psi}^T\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^T\mathbf{v}\}$$
- If $E\{\boldsymbol{\Psi}^T\mathbf{v}\} = \mathbf{0}$, the estimate is bias free:
  - In $\boldsymbol{\Psi}^T\mathbf{v}$ only products of the types $v(i)y(i-c)$ and $v(i)u(i-c)$ can be found, where i is an arbitrary integer, c is an arbitrary positive integer
  - The estimate $\hat{\boldsymbol{\theta}}$ is therefore bias free if the noise $v(k)$ is uncorrelated with the past input and output process values and either noise or input and output are zero-mean:
    - current noise not correlated with past inputs $\rightarrow$ not a problem
    - current noise not correlated with past outputs $\rightarrow$ could be a problem since:
      - $v(k)$ always correlated with $y(k)$, $v(k-1)$ always correlated with $y(k-1)$, etc.
      - $v(k)$ not correlated with $y(k-1)$ only if $v(k)$ not correlated with $v(k-1)$ $\rightarrow$ this is true if $v(k)$ is white noise

Bias, consistency and convergence of estimates

- The parameter estimate $\widehat{\boldsymbol{\theta}}$ is bias free if:
  1. the process output noise $n(k)$ is obtained by filtering white noise $v(k)$ through a noise filter $\frac{1}{A(z^{-1})}$, and
  2. either noise $v(k)$ or process input $u(k)$ is zero-mean
- When the above conditions are met:
  - $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta} \Rightarrow \mathbf{e} = \mathbf{v}$ which means $e(k) = v(k)$ for all k from the interval
- Generalised error $e(k)$ therefore converges to noise $v(k)$:
  - In the schematic representation of the error model, $e(k)$ is obtained by filtering $v(k)$ through $1/A(z^{-1})$ and then feeding the result to $\hat{A}(z^{-1})$
  - Due to bias-free estimate $\hat{A}(z^{-1}) = A(z^{-1})$ the transfer function from $v(k)$ to $e(k)$ becomes unity: $e(k) = v(k)$
  - Two signal paths from $u(k)$ to $e(k)$ cancel each other if $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$ which means that $e(k)$ is not controllable from $u(k)$ (this is true only if the estimated parameters are exactly the same as the true ones)

Bias, consistency and convergence of estimates

- What about the covariance matrix of the parameter estimate error?

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right]=E\left\{\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right]\times\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right]^{T}\right\}=E\{[\boldsymbol{\Psi}^{T}\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^{T}\mathbf{v}\mathbf{v}^{T}\boldsymbol{\Psi}[\boldsymbol{\Psi}^{T}\boldsymbol{\Psi}]^{-1}\}$$

- White noise $v(k)$ is not correlated with $\boldsymbol{\Psi}$, and $e(k)=v(k)$:

$$E\{\mathbf{v}\mathbf{v}^{T}\}=E\{\mathbf{e}\mathbf{e}^{T}\}=\sigma_{e}^{2}\mathbf{I}$$

- Covariance matrix of the parameter estimate error becomes:

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right]=\sigma_{e}^{2}[\boldsymbol{\Psi}^{T}\boldsymbol{\Psi}]^{-1}=\frac{1}{N}\sigma_{e}^{2}[\frac{1}{N}\boldsymbol{\Psi}^{T}\boldsymbol{\Psi}]^{-1}$$

- Elements on the main diagonal of the covariance matrix are variances (squares of standard deviations) of the corresponding parameter estimate errors
- Matrix $\boldsymbol{\Psi}^{T}\boldsymbol{\Psi}$ is the correlation matrix multiplied by N:
  - Inverse of $\boldsymbol{\Psi}^{T}\boldsymbol{\Psi}$ is the inverse of the correlation matrix divided by N
    $\Rightarrow$ converges to 0 as $N\to\infty$ which means $\mathrm{cov}\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right]\to\mathbf{0}$
    $\Rightarrow$ **the estimate $\hat{\theta}$ is consistent in the mean square**

Recursive least squares (RLS) method

- The least squares LS method just treated provides parameter estimates after a batch of previously collected data has been processed – it belongs to a group of offline methods with non-recursive estimation algorithm and direct (one-step) processing
- If parameters need to be estimated online, the estimates are updated as new measurements from the process arrive
- Each new pair of input-output measurements (at time $k+1$) adds some information that is mathematically reflected in an additional equation or alternatively in extending vector $\mathbf{y}$ with $y(k+1)$ and extending matrix $\boldsymbol{\Psi}$ with a row $\boldsymbol{\psi}^T(k+1)$
- If a non-recursive method were applied, the product $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ would have to be inverted and multiplied by $\boldsymbol{\Psi}^T\mathbf{y}$ in each sampling time (very time consuming)
- The equations will therefore be put to a recursive form

Recursive least squares (RLS) method

- Recursive least squares method is **mathematically equivalent to the non-recursive least squares method**, but the updated estimate is obtained in each sampling time based on the previous estimate (without calculation of the matrix inverse)
- Matrices and vectors change with time and will be denoted as discrete-time signals (with argument k)
- Non-recursive parameter estimation therefore reads:

$$\widehat{\boldsymbol{\theta}}(k) = [\boldsymbol{\Psi}^T(k)\boldsymbol{\Psi}(k)]^{-1}\boldsymbol{\Psi}^T(k)\mathbf{y}(k)$$

- A new matrix is introduced:

$$\mathbf{P}(k) = [\boldsymbol{\Psi}^T(k)\boldsymbol{\Psi}(k)]^{-1}$$

- The extension of vectors and matrices at time $k+1$:

$$\mathbf{y}(k+1) = \begin{bmatrix} \mathbf{y}(k) \\ y(k+1) \end{bmatrix} \qquad \boldsymbol{\Psi}(k+1) = \begin{bmatrix} \boldsymbol{\Psi}(k) \\ \boldsymbol{\psi}^T(k+1) \end{bmatrix}$$

$$\boldsymbol{\psi}^T(k+1) = [-y(k) \quad \dots \quad -y(k-n+1) \quad u(k-d) \quad \dots \quad u(k-n-d+1)]$$

Recursive least squares (RLS) method

- It is very easy to obtain recursive form for $\mathbf{P}^{-1}$:

$$\mathbf{P}^{-1}(k+1) = \mathbf{P}^{-1}(k) + \boldsymbol{\psi}(k+1)\boldsymbol{\psi}^T(k+1), \quad \mathbf{P}^{-1}(0) = \mathbf{0}$$

- Recursive equation for $\hat{\boldsymbol{\theta}}$ is obtained after a short derivation:

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \mathbf{P}(k+1)\boldsymbol{\psi}(k+1)\left[y(k+1) - \boldsymbol{\psi}^T(k+1)\hat{\boldsymbol{\theta}}(k)\right]$$

updated estimate = last estimate + correction gain [new measurement − predicted measurement based on last estimate]

  - An initial condition $\hat{\boldsymbol{\theta}}(0)$ is needed for solving this difference equation
- $\mathbf{P}(k+1)$ is needed in calculation of $\hat{\boldsymbol{\theta}}(k+1)$ but recursive form is only given for its inverse $\mathbf{P}^{-1}$
- Applying *matrix inversion lemma* (Woodbury matrix id.) results in:

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \frac{\mathbf{P}(k)\boldsymbol{\psi}(k+1)\boldsymbol{\psi}^T(k+1)\mathbf{P}(k)}{1 + \boldsymbol{\psi}^T(k+1)\mathbf{P}(k)\boldsymbol{\psi}(k+1)}$$

  - An initial condition $\mathbf{P}(0)$ needed for solving this difference eq.
  - Initial condition should be „large“; usual choice: $\mathbf{P}(0) = \alpha\mathbf{I} \quad \alpha \gg 1$

Method of weighted least squares (WLS)

- In least squares parameter estimation all the data are given the same weight when calculating cost function V
- If the level of confidence to measurements varies with time, the equation errors $e(k)$ are given different weights
- The cost function V changes from $V = \mathbf{e}^T\mathbf{e}$ to:

$$V = \mathbf{e}^T\mathbf{W}\mathbf{e}$$

  - $\mathbf{W}$ is a (diagonal) matrix of individual weights:

$$\mathbf{W} = \begin{bmatrix} w(k-N+1) & 0 & 0 & \cdots & 0 \\ 0 & w(k-N+2) & 0 & \cdots & 0 \\ 0 & 0 & w(k-N+3) & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & w(k) \end{bmatrix}$$

  - From theoretical point of view $\mathbf{W}$ can be an arbitrary positive definite matrix but the interpretation of weights is then more difficult

Method of weighted least squares (WLS)

- The parameter estimate is obtained following the same procedure as in the case of least squares parameter estimation:

$$\widehat{\boldsymbol{\theta}} = [\boldsymbol{\Psi}^T \mathbf{W} \boldsymbol{\Psi}]^{-1} \boldsymbol{\Psi}^T \mathbf{W} \mathbf{y}$$

- The conditions for bias free and consistency also remain the same
- Covariance matrix of the parameter estimate error is:

$$\operatorname{cov}\left[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = E\{[\boldsymbol{\Psi}^T \mathbf{W} \boldsymbol{\Psi}]^{-1} \boldsymbol{\Psi}^T \mathbf{W} \mathbf{v}\mathbf{v}^T \mathbf{W} \boldsymbol{\Psi} [\boldsymbol{\Psi}^T \mathbf{W} \boldsymbol{\Psi}]^{-1}\}$$

- The smallest variance is obtained when $\mathbf{W}$ is chosen as inverse of the covariance matrix of the noise $v(k)$:

$$\mathbf{W}^{-1} = E\{\mathbf{v}\mathbf{v}^T\}$$

  - Estimates with minimum variance are also termed Markov estimates
  - The covariance matrix of the noise is in general not known a priori
  - If $v(k)$ is a white noise, its covariance matrix is unity matrix (multiplied by a scalar), and WLS parameter estimation equation reduces to LS parameter estimation

Exponential forgetting (EF)

- Forgetting means that equation errors based on old data have smaller weights than the ones based on recent measurements
- This approach is useful in the case of time-varying systems
- Exponential forgetting is a special case of weighting:
  - When the data are fresh (at time k), the weight is $w(k)$
  - In the following sampling time $(k+1)$, the weight of this equation reduces to $\lambda w(k)$ where $\lambda < 1$
  - At sampling time $(k+i)$ the weight becomes $\lambda^i w(k)$
  - The weight of an individual equation is multiplied by a factor $\lambda < 1$ in each sampling time $\rightarrow$ the weight goes to 0 exponentially with time $\rightarrow$ the influence of a measurement gradually diminishes
  - The choice $\lambda = 1$ means that there is no forgetting $\rightarrow$ WLS
  - Interval of parameter λ:

$$0 < \lambda \leq 1 \text{ (usually } 0{,}95 \leq \lambda \leq 0{,}999)$$

Recursive WLS (RWLS) with exponential forgetting

- Forgetting can be time-varying in general $\rightarrow \lambda(k)$
- Weighting matrix can be given in recursive form:

$$\mathbf{W}(k+1) = \begin{bmatrix} \lambda(k+1)\mathbf{W}(k) & \mathbf{0} \\ \mathbf{0}^T & w(k+1) \end{bmatrix}$$

- Matrix $\mathbf{P}(k)$ is defined slightly different than with RLS:

$$\mathbf{P}(k) = [\boldsymbol{\Psi}^T(k)\mathbf{W}(k)\boldsymbol{\Psi}(k)]^{-1}$$

- Parameter estimates are given by:

$$\widehat{\boldsymbol{\theta}}(k) = \mathbf{P}(k)\boldsymbol{\Psi}^T(k)\mathbf{W}(k)\mathbf{y}(k)$$

- Recursive equation for $\mathbf{P}^{-1}$ has to include (varying) weights of measurements and (varying) exponential forgetting factors:

$$\mathbf{P}^{-1}(k+1) = \lambda(k+1)\mathbf{P}^{-1}(k) + w(k+1)\boldsymbol{\psi}(k+1)\boldsymbol{\psi}^T(k+1)$$

- By using matrix inversion lemma we obtain:

$$\mathbf{P}(k+1) = \left(\mathbf{P}(k) - \frac{w(k+1)\mathbf{P}(k)\boldsymbol{\psi}(k+1)\boldsymbol{\psi}^T(k+1)\mathbf{P}(k)}{\lambda(k+1) + w(k+1)\boldsymbol{\psi}^T(k+1)\mathbf{P}(k)\boldsymbol{\psi}(k+1)}\right)\frac{1}{\lambda(k+1)}$$

RWLS with exponential forgetting (RWLSEF)

- Recursive parameter estimation equation in the case of RWLSEF:
$$\widehat{\boldsymbol{\theta}}(k+1) = \widehat{\boldsymbol{\theta}}(k) + w(k+1)\mathbf{P}(k+1)\boldsymbol{\psi}(k+1)\left[y(k+1) - \boldsymbol{\psi}^T(k+1)\widehat{\boldsymbol{\theta}}(k)\right]$$
- Comparing the above update equation with the one obtained by RLS, only an extra factor $w(k+1)$ is added to the correction gain
- Update equation for $\mathbf{P}$ hides a potential danger:
  - If $\lambda < 1$ and excitation is absent ($\boldsymbol{\psi}(k+1) = \mathbf{0}$):
    - matrix $\mathbf{P}^{-1}(k)$ converges to 0, or equivalently its inverse converges to ∞
    - small difference between new measurement and the predicted measurement based on last estimate is detected due to e.g. noise
    - this small difference is amplified by high correction gain ($w\mathbf{P}\boldsymbol{\psi}$)
    - extremely large (wrong) parameter update
  - Briefly: we say that algorithms first „falls asleep“ and then „blows up“
  - Large $\mathbf{P}(k)$ also means large parameter variances (bad precision)
  - To prevent this – $\lambda(k)$ should adapt to information content in $\boldsymbol{\psi}$

Parameter estimation for time-varying processes

- If time-varying processes are estimated, continuous changes of parameters should be allowed for $\rightarrow$ matrix $\mathbf{P}(k)$ which is a part of correction gain should be prevented from going to 0
- Two principal approaches that achieve this goal:
  - Forgetting – its effect is that the elements of $\mathbf{P}(k)$ are slightly increased in each sampling time – convergence to 0 not possible
  - Covariance resetting:
    - From time to time $\mathbf{P}(k)$ is set to some constant matrix
    - If process parameters change in discrete times it would be ideal if $\mathbf{P}(k)$ is reset in these moments; these moments are unknown in advance, so:
      - matrix $\mathbf{P}(k)$ is reset periodically or
      - matrix $\mathbf{P}(k)$ is reset when its „size“ (usually trace) drops below a threshold
- Both methods have pros and cons:
  - Resetting useful when the changes are in the form of periodic jumps
  - Forgetting useful when changes are drift-like

Unknown steady-state values

- Measurements $u(k)$ and $y(k)$ that constitute vector $\boldsymbol{\psi}(k)$ are not original readings from sensors – they are deviations from steady-state values that are usually referred to as operating point:

$$u(k) = U(k) - U_{00}$$
$$y(k) = Y(k) - Y_{00}$$

  - $U(k)$ and $Y(k)$ are absolute values of input and output, respectively
  - U_{00} and Y_{00} are steady-state values of input and output, respectively
  - Input $U(k) = U_{00}$ results in the output $Y(k)$ with steady-state Y_{00}
- A simple approach for determination of steady-state values is to require zero-mean long-time average of $u(k)$ and $y(k)$:

$$\sum_{k=0}^{N} u(k) = 0 \quad \text{and} \quad \sum_{k=0}^{N} y(k) = 0$$

  - Or equivalently:

$$U_{00} = \frac{1}{N+1}\sum_{k=0}^{N} U(k) \quad \text{and} \quad Y_{00} = \frac{1}{N+1}\sum_{k=0}^{N} Y(k)$$

Unknown steady-state values

- Noise $n(k)$ is an integral part of the output
- If mean value of the noise is not 0, only unknown steady state of the output will be affected, and therefore we can assume that mean value (mathematical expectation) of the noise is 0:

$$E\{n(k)\} = 0$$

- Operating point can be determined as $Y_{00} = \frac{1}{N+1}\sum_{k=0}^{N} Y(k)$, but this approach is only possible when parameters are estimated offline
- Online identification can be combined with the following approaches to cope with unknown steady-state values:
  - **Operating point is estimated** with process parameters
  - The effect of operating point is **cancelled with differencing**

Estimation of operating point

- Introducing $u(k) = U(k) - U_{00}$ and $y(k) = Y(k) - Y_{00}$ into process difference equation yields:

$$Y(k) = -a_1 Y(k-1) - \cdots - a_n Y(k-n) + b_1 U(k-d-1) + \cdots + b_n U(k-d-n) + \\ + \boxed{Y_{00} + a_1 Y_{00} + \cdots + a_n Y_{00} - b_1 U_{00} - \cdots - b_n U_{00}} + v(k)$$

- All terms due to steady state values can be combined:

$$K = Y_{00} + a_1 Y_{00} + \cdots + a_n Y_{00} - b_1 U_{00} - \cdots - b_n U_{00}$$

- Parameter K is added to the vector of unknown parameters:

$$\boldsymbol{\theta}^T = [a_1, \ldots, a_n, b_1, \ldots, b_n, K]$$

$$\boldsymbol{\psi}^T(k) = [-Y(k-1), \ldots, -Y(k-n), U(k-d-1), \ldots, U(k-d-n), 1]$$

  - Regressors $\boldsymbol{\psi}$ now include absolute measurements (not the deviations), and element 1 is included at the end

- The process equation takes the following form:

$$Y(k) = \boldsymbol{\psi}^T(k)\boldsymbol{\theta} + v(k)$$

  - The equations has standard form, and the problem solution is also a standard one

Cancelling the steady-state values by differencing

- Steady state values U_{00} and Y_{00} in

$$u(k) = U(k) - U_{00} \quad \text{and} \quad y(k) = Y(k) - Y_{00}$$

assumed constant $\rightarrow$ the differences of absolute signals are the same as the differences of deviations:

$$\Delta U(k) = U(k) - U(k-1) = u(k) - u(k-1) = \Delta u(k)$$
$$\Delta Y(k) = Y(k) - Y(k-1) = y(k) - y(k-1) = \Delta y(k)$$

- Recall the process difference equation:

$$y(k) = -a_1 y(k-1) - \cdots - a_n y(k-n) + b_1 u(k-d-1) + \cdots + b_n u(k-d-n) + v(k)$$

- If the above equation is rewritten for the time $k-1$ and subtracted from the original equation for the time k, we get:

$$\Delta y(k) = -a_1 \Delta y(k-1) - a_2 \Delta y(k-2) - \cdots - a_n \Delta y(k-n) + b_1 \Delta u(k-d-1) + \cdots + b_n \Delta u(k-d-n) + \boxed{v(k) - v(k-1)}$$

Cancelling the steady-state values by differencing

- Due to $\Delta Y(k) = \Delta y(k)$ and $\Delta U(k) = \Delta u(k)$ the equation can be rewritten with absolute signal differences:

$$\Delta Y(k) = -a_1 \Delta Y(k-1) - a_2 \Delta Y(k-2) - \cdots - a_n \Delta Y(k-n) + b_1 \Delta U(k-d-1) + \cdots + b_n \Delta U(k-d-n) + v(k) - v(k-1)$$

- The estimation equation takes the following form:

$$\Delta Y(k) = \boldsymbol{\psi}^T(k)\boldsymbol{\theta} + v(k) - v(k-1)$$

$$\boldsymbol{\psi}^T(k) = [-\Delta Y(k-1), \ldots, -\Delta Y(k-n), \Delta U(k-d-1), \ldots, \Delta U(k-d-n)]$$

- The estimation equation is very similar to the standard one (only $v(k) - v(k-1)$ takes the role of $v(k)$) $\rightarrow$ standard solution
- What is different from the standard case?
  - Differencing amplifies high-frequency noise, or equivalently
  - the estimate is bias-free if $v(k) - v(k-1)$ is a white noise

    $\Leftrightarrow v(k)$ is obtained by filtering white noise through $\frac{1}{1-z^{-1}}$

    $\Leftrightarrow n(k)$ obtained by filtering white noise through $\frac{1}{(1-z^{-1})A(z^{-1})}$

Unknown steady-state values

- Pros and cons of both **on-line** approaches:
  - Extra estimated parameter $\widehat{K}$:
    + less sensitive to noise
    + estimate $\widehat{K}$ gives mathematical relation between U_{00} and Y_{00}
    – an extra estimated parameter → slower convergence
  - Differencing:
    + number of estimated parameters the same as in LS approaches
    – differencing amplifies (high-frequency) noise and thus affects convergence
- A very simple **off-line** approach to determine Y_{00} is to „measure“ it directly (before start of dynamical process identification):
  - input U_{00} is applied to the input
  - wait until the output settles
  - Y_{00} is the average of the following N_{00} samples (input is still U_{00})

Numerical problems

- Numerical problems arise if parameter estimation is applied on an ill-conditioned system of equations
- Parameter identification requires that the order of the process and time delay are known in advance:
  - If exact order is not known, a higher order can be selected as it always includes a lower one $\rightarrow$ this leads to extra regressors that do not affect the output – the corresponding „true“ parameters are 0
  - Under-estimated delay also leads to ill-conditioned estimation problem (if d is smaller than the actual delay d_0, the parameters $b_1, \ldots, b_{d_0-d}$ are redundant – „true“ value is 0 again)
  - The consequence of redundant parameters is ill-conditioned system:
    - Pole-zero pairs that lie close together are estimated
    - They have negligible contribution to the input-output process behaviour
    - If they move drastically, the output change is not significant
    - If they move to unstable region, this DOES seriously affect the model

Numerical problems

- Ill-conditioned system of equations (or linear dependence of vectors in $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$) can also arise if:
  - sampling time is too small (too fast sampling)
  - steady-state of signals is very large w.r.t. signal deviations
- Ill-conditioned problem (or ill-conditioned $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$) is reflected as:
  - Small changes (errors) in measurements lead to large changes in estimated parameters
  - Estimated parameters that significantly deviate from true ones result in small equation errors (usually output or generalised error)
- This phenomenon is quantified with a condition number κ that will be derived for the case of a linear system of equations:

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{b} \qquad \mathbf{A} = \boldsymbol{\Psi}^T\boldsymbol{\Psi}, \mathbf{b} = \boldsymbol{\Psi}^T\mathbf{y}$$

- In case of an error $\Delta\mathbf{b}$ on $\mathbf{b}$, estimated $\boldsymbol{\theta}$ also changes for $\Delta\boldsymbol{\theta}$:

$$\mathbf{A}(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) = \mathbf{b} + \Delta\mathbf{b} \quad \Rightarrow \quad \Delta\boldsymbol{\theta} = \mathbf{A}^{-1}\Delta\mathbf{b}$$

Numerical problems

- The norms of vectors $\mathbf{b} = \mathbf{A\theta}$ and $\Delta\boldsymbol{\theta} = \mathbf{A}^{-1}\Delta\mathbf{b}$ will be analysed:

$$\|\mathbf{b}\| = \|\mathbf{A\theta}\| \leq \|\mathbf{A}\| \cdot \|\boldsymbol{\theta}\|$$
$$\|\Delta\boldsymbol{\theta}\| = \|\mathbf{A}^{-1}\Delta\mathbf{b}\| \leq \|\mathbf{A}^{-1}\| \cdot \|\Delta\mathbf{b}\|$$

- Rearranging the non-equalities we get:

$$\left.\begin{aligned} \frac{1}{\|\boldsymbol{\theta}\|} &\leq \frac{\|\mathbf{A}\|}{\|\mathbf{b}\|} \\ \|\Delta\boldsymbol{\theta}\| &\leq \|\mathbf{A}^{-1}\| \cdot \|\Delta\mathbf{b}\| \end{aligned}\right\} \Rightarrow \frac{\|\Delta\boldsymbol{\theta}\|}{\|\boldsymbol{\theta}\|} \leq \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\| \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}$$

- Relative error in estimated parameters $\boldsymbol{\theta}$ is obtained by amplifying the relative error in $\mathbf{b}$ with **condition number** of a matrix $\kappa(\mathbf{A})$:

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\| \geq 1$$

- Small condition number $\rightarrow$ small sensitivity of estimated parameters to measurements errors in $\mathbf{b}$
- Large condition number $\rightarrow$ large sensitivity of estimated parameters to measurements errors in $\mathbf{b}$ – ill-conditioned system

Numerical problems

- What is $\|\mathbf{A}\|$ anyway?
  - Matrix norm is always induced by the choice of a vector norm
  - If an Euclidean vector norm (Pitagora length) is chosen, the corresponding or induced matrix norm is the so called spectral norm:

$$\|\mathbf{A}\| = \max \sigma_i(\mathbf{A})$$

- What is $\sigma_i(\mathbf{A})$?
  - $\sigma_i(\mathbf{A})$ are matrix singular values – always real and positive
  - A square matrix (dim. $n \times n$) has n singular values:

$$\sigma_i(\mathbf{A}) = \sqrt{\lambda_i(\mathbf{A}^T\mathbf{A})} = \sqrt{\lambda_i(\mathbf{A}\mathbf{A}^T)} \quad i = 1,2,\dots n$$

- Norm $\|\mathbf{A}\|$ is therefore the largest singular value $\sigma_{\max}(\mathbf{A})$, and norm $\|\mathbf{A}^{-1}\|$ is the inverse of the smallest singular value $\frac{1}{\sigma_{\min}(\mathbf{A})}$
- Condition number κ is the product $\kappa(\mathbf{A}) = \frac{\sigma_{max}}{\sigma_{min}} \geq 1$

Method of Extended Least Squares (ELS)

- One of the conditions for bias-free estimates with LS method: $n(k)$ should be obtained filtering white noise through $G_n(z) = \frac{1}{A(z^{-1})}$
- This is a strict requirement – very rarely fulfilled in practice
- To relax this condition a more general noise filter is assumed:

$$G_n(z) = \frac{D(z^{-1})}{A(z^{-1})} = \frac{1 + d_1 z^{-1} + \cdots + d_n z^{-n}}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}}$$

- Equation of the process with disturbance becomes:

$$y(z) = \frac{B(z^{-1})}{A(z^{-1})} z^{-d} u(z) + \frac{D(z^{-1})}{A(z^{-1})} v(z)$$

$$A(z^{-1}) y(z) = B(z^{-1}) z^{-d} u(z) + D(z^{-1}) v(z)$$

- Equation in the time domain:

$$y(k) = -a_1 y(k-1) - a_2 y(k-2) - \cdots - a_n y(k-n)$$
$$+ b_1 u(k-d-1) + \cdots + b_n u(k-d-n)$$
$$+ v(k) + d_1 v(k-1) + \cdots + d_n v(k-n)$$

Method of Extended Least Squares (ELS)

- Extended vectors are defined as:

$$\boldsymbol{\psi}_0^T(k) = [-y(k-1), \dots, -y(k-n), u(k-d-1), \dots, u(k-d-n), v(k-1), \dots, v(k-n)]$$

$$\boldsymbol{\theta}^T = [a_1, \dots, a_n, b_1, \dots, b_n, d_1, \dots d_n]$$

- Process equation:

$$y(k) = \boldsymbol{\psi}_0^T(k)\boldsymbol{\theta} + v(k)$$

- If noise $v(k)$ and its delayed samples $v(k-1), \dots, v(k-n)$ were known, LS estimation method could be used for estimation of extended vector of parameters:

$$\hat{\boldsymbol{\theta}}^T = [\hat{a}_1, \dots, \hat{a}_n, \hat{b}_1, \dots, \hat{b}_n, \hat{d}_1, \dots, \hat{d}_n]$$

- Noise $v(k)$ cannot be known but equation error $e(k)$ is known
- In case of bias-free estimation it holds $v(k) = e(k) \rightarrow$ in vector $\boldsymbol{\psi}_0^T(k)$ delayed noise is substituted with delayed errors:

$$\boldsymbol{\psi}^T(k) = [-y(k-1), \dots, -y(k-n), u(k-d-1), \dots, u(k-d-n), e(k-1), \dots, e(k-n)]$$

Method of Extended Least Squares (ELS)

- Now any version of recursive LS can be used but matrix $\mathbf{P}(k)$ and vectors $\widehat{\boldsymbol{\theta}}(k)$ and $\boldsymbol{\psi}(k)$ are of larger dimensions $(3n \times 3n,\ 3n \times 1)$
- With non-recursive LS the definition of $e(k)$ is clear:

$$e(k) = y(k) - \boldsymbol{\psi}^T(k)\widehat{\boldsymbol{\theta}}$$

- With recursive LS, the question arises which $\widehat{\boldsymbol{\theta}}$ to use:
  - If the estimate from the previous sample time is used, **a priori error** $e(k)$ is obtained:

$$e(k) = y(k) - \boldsymbol{\psi}^T(k)\widehat{\boldsymbol{\theta}}(k-1)$$

  - If the estimate from the current sample time is used, **a posteriori error** $\varepsilon(k)$ is obtained:

$$\varepsilon(k) = y(k) - \boldsymbol{\psi}^T(k)\widehat{\boldsymbol{\theta}}(k)$$

- Two versions of error $\rightarrow$ two versions of extended least squares – they only differ in the last n elements in the vector $\boldsymbol{\psi}$ (e or ε)

Method of Extended Least Squares (ELS)

- It turns out that both versions of error calculation have similar properties when used in ELS, and therefore a priori error $e(k)$ is used more often because it is needed for calculation of $\widehat{\boldsymbol{\theta}}(k)$
- Algorithm in the case of a priori error is therefore a for-loop over k:

$$\boldsymbol{\psi}^T(k) = [-y(k-1), \dots, -y(k-n), u(k-d-1), \dots, u(k-d-n), e(k-1), \dots, e(k-n)]$$

$$\mathbf{P}(k) = \left(\mathbf{P}(k-1) - \frac{w(k)\mathbf{P}(k-1)\boldsymbol{\psi}(k)\boldsymbol{\psi}^T(k)\mathbf{P}(k-1)}{\lambda(k) + w(k)\boldsymbol{\psi}^T(k)\mathbf{P}(k-1)\boldsymbol{\psi}(k)}\right)\frac{1}{\lambda(k)}$$

$$e(k) = y(k) - \boldsymbol{\psi}^T(k)\widehat{\boldsymbol{\theta}}(k-1)$$

$$\widehat{\boldsymbol{\theta}}(k) = \widehat{\boldsymbol{\theta}}(k-1) + w(k)\mathbf{P}(k)\boldsymbol{\psi}(k)\underbrace{\left[y(k) - \boldsymbol{\psi}^T(k)\widehat{\boldsymbol{\theta}}(k-1)\right]}_{e(k)}$$

- A priori error in the explicit form:

$$e(k) = y(k) + \hat{a}_1 y(k-1) + \cdots + \hat{a}_n y(k-n)$$
$$-\hat{b}_1 u(k-d-1) - \cdots - \hat{b}_n u(k-d-n) - \hat{d}_1 e(k-1) - \cdots - \hat{d}_n e(k-n)$$

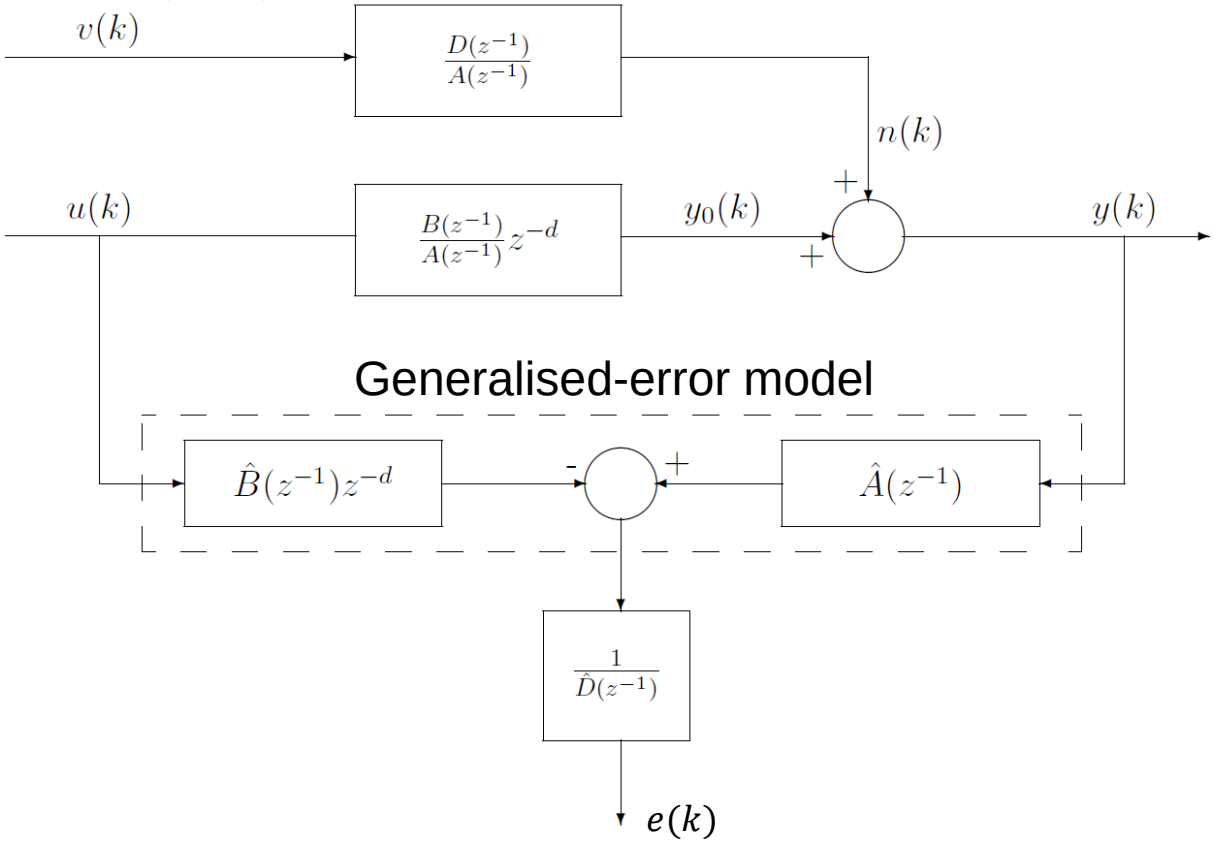
- Its z transform:

$$e(z) = \hat{A}(z^{-1})y(z) - \hat{B}(z^{-1})z^{-d}u(z) + [1 - \hat{D}(z^{-1})]e(z)$$

Method of Extended Least Squares (ELS)

- From the previous equation $e(z)$ can be derived:

$$e(z) = \frac{1}{\hat{D}(z^{-1})}\left(\hat{A}(z^{-1})y(z) - \hat{B}(z^{-1})z^{-d}u(z)\right)$$



Method of Extended Least Squares (ELS)

- This figure illustrates the main idea of ELS:
  - If discrete white noise $v(k)$ is filtered through $G_n(z)$, variance of output noise $n(k)$ is obtained from input noise variance σ_v^2 and samples of the filter impulse response $g_n(k) = \mathcal{Z}^{-1}\{G_n(z)\}$:
    $$\sigma_n^2 = \sigma_v^2(g_n^2(0) + g_n^2(1) + g_n^2(2) + \cdots)$$
  - In case of monic polynomials (leading coefficient is 1) in $G_n(z)$:
    - The impulse response always starts with $g_n(0) = 1$
    - The lowest variance σ_n^2 is obtained if $g_n(k)|_{k>0} = 0 \Leftrightarrow G_n(z) = 1$
  - See figure: $e(k)$ is obtained by filtering $v(k)$ through filter with monic numerator $D(z^{-1})\hat{A}\,(z^{-1})$ and monic denominator $A(z^{-1})\widehat{D}\,(z^{-1})$
  - Least squares approach tries to minimise the variance of $e(k) \rightarrow$ filter from $v(k)$ to $e(k)$ converges to $1 \rightarrow e(k)$ converges to $v(k)$
  - Numerator $D(z^{-1})$ in the noise filter $G_n(z) \rightarrow$ the error obtained in the case of classical LS is additionally filtered by $\frac{1}{\widehat{D}\,(z^{-1})}$

Method of Extended Least Squares (ELS)

- Advantage of ELS if compared to the classical LS approach:
  - Noise filter $\frac{D(z^{-1})}{A(z^{-1})}$ is much more general than $\frac{1}{A(z^{-1})}$ that was assumed with LS method:
    - The class of noise signals $n(k)$ that satisfy the bias-free conditions is much broader than in the case of LS
    - The LS condition is included in the ELS conditions: $D(z^{-1}) = 1$
- Weaknesses of ELS when compared to LS:
  - Higher number of estimated parameters ($3n$ versus $2n$) $\rightarrow$ slower convergence, especially the convergence of $\hat{d}_i$ is very slow
  - A more critical problem is stability of the algorithm:
    - Error signal $e(k)$ or $\varepsilon(k)$ that is obtained in step k is used in vector $\boldsymbol{\psi}(k+1)$ that is used to calculate error in step $k+1$
    - This introduces a feedback into the estimation problem
    - Feedback can potentially destabilise the system – estimation problem

Stability of the ELS method

- ELS method is locally stable in the neighbourhood of $\tilde{\boldsymbol{\theta}} = \mathbf{0}$ if:

$$\mathrm{Re}\left[\frac{1}{D(e^{j\omega})}\right] > 0 \qquad \text{for all real } \omega$$

  - The neighbourhood of $\tilde{\boldsymbol{\theta}} = \mathbf{0}$ means that the estimates should always stay in the vicinity of true values – the initial condition should also lie „close“ to the true values
- ELS method is globally stable if:

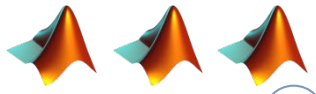
$$\mathrm{Re}\left[\frac{1}{D(e^{j\omega})} - \frac{\lambda}{2}\right] > 0 \qquad \text{for all real } \omega$$

  - This condition is more strict and harder to fulfil
- Stability of an identification method – parameter estimates do not diverge and therefore stay bounded
- Stability does not imply bias-free and consistent parameter estimates

Bias and consistency of the ELS method

- Parameter estimates converge to true values of parameters if stability conditions are fulfilled and the following statements hold:
  - Matrix $\mathbf{P}(k)$ converges to 0
  - Noise $v(k)$ on the input of the noise filter is a white noise
- The first condition is fulfilled if $\boldsymbol{\psi}(k)$ (or $\boldsymbol{\psi}_{\varepsilon}(k)$ in the case of an a posteriori error) is a persistently exciting vector:
  - No component converges to 0
  - The components are linearly independent
- If any of the conditions is not fulfilled, the estimates are biased
- Classical LS with EF where $D(e^{j\omega})=1 \rightarrow$ locally and globally stable:

$$\mathrm{Re}\left[\frac{1}{D(e^{j\omega})}-\frac{\lambda}{2}\right]>0, \qquad 0<\lambda\leq 1$$



Maximum likelihood

- Maximum likelihood is a statistical method
- Åström and Bohlin used this approach for estimation of parameters of dynamical systems as early as in 1965
- The idea is that those parameters are taken as the estimates $\hat{\boldsymbol{\theta}}$, which make the measurement most likely to occur:
  - Select the parameter estimate $\hat{\boldsymbol{\theta}}$ that will maximise the occurrence of the measurements taking into account the class of process models
  - The likelihood function $p(y|\boldsymbol{\theta})$ has to be defined
  - The parameter vector $\boldsymbol{\theta}$ that maximizes the likelihood function $p(y|\boldsymbol{\theta})$ has to be found $\rightarrow$ this is the solution of identification $\hat{\boldsymbol{\theta}}$
- The algorithm is much simplified if normally distributed and statistically independent signals are assumed (the same assumptions on disturbances as in the case of ELS)

Maximum likelihood

- There exist non-recursive and recursive version:
  - Note that non-recursive method is still iterative due to nonlinear relation between parameters and signal error
- The derivation of the algorithm is quite complicated
- However, implementation is simple; first define filtered signals:

$$y_f(k) = y(k) - \hat{d}_1 y_f(k-1) - \cdots - \hat{d}_n y_f(k-n)$$
$$u_f(k) = u(k) - \hat{d}_1 u_f(k-1) - \cdots - \hat{d}_n u_f(k-n)$$
$$e_f(k) = e(k) - \hat{d}_1 e_f(k-1) - \cdots - \hat{d}_n e_f(k-n)$$

$$\boldsymbol{\varphi}^T(k) = \left[-y_f(k-1), \ldots, -y_f(k-n), u_f(k-d-1), \ldots, u_f(k-d-n), e_f(k-1), \ldots, e_f(k-n)\right]$$

  - For the generation the estimates of noise filter numerator are used
- Recursive formulas for $\hat{\boldsymbol{\theta}}$ and $\mathbf{P}$ (5x $\boldsymbol{\varphi}$ instead of $\boldsymbol{\psi}$, 1 $\boldsymbol{\psi}$ remains):

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + w(k+1)\mathbf{P}(k+1)\boldsymbol{\varphi}(k+1)\left[y(k+1) - \boldsymbol{\psi}^T(k+1)\hat{\boldsymbol{\theta}}(k)\right]$$

$$\mathbf{P}(k+1) = \left(\mathbf{P}(k) - \frac{w(k+1)\mathbf{P}(k)\boldsymbol{\varphi}(k+1)\boldsymbol{\varphi}^T(k+1)\mathbf{P}(k)}{\lambda(k+1) + w(k+1)\boldsymbol{\varphi}^T(k+1)\mathbf{P}(k)\boldsymbol{\varphi}(k+1)}\right)\frac{1}{\lambda(k+1)}$$

Instrumental variables

- Recall: classical LS parameter estimation methods gives bias-free estimates if cross-correlation functions $\phi_{yv}(k)$ are zero for positive values of argument:

$$\phi_{yv}(k) \qquad \text{for } k = 1,2 \ldots n$$

  - Satisfied only with a special form of the noise filter $1/A(z^{-1})$
- Instrumental variables method suppresses the bias by replacing measured output values y (corrupted with noise) with the so-called instrumental variables x:
  - instrumental variables x not correlated with disturbance v
  - instrumental variables x of course highly correlated with y
- The method provides unbiased estimates with an arbitrary form of the noise filter (a more general noise is allowed), but only local stability is retained (the price we pay is that stability is weaker)

Instrumental variables

- The derivation:
  - The LS parameter estimation formula can be derived informally by pre-multiplying equation $\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\theta} = \mathbf{v}$ with $\boldsymbol{\Psi}^T$ and expressing $\boldsymbol{\theta}$:
$$\boldsymbol{\theta} = [\boldsymbol{\Psi}^T\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^T\mathbf{y} - [\boldsymbol{\Psi}^T\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^T\mathbf{v}$$
    - Estimate $\widehat{\boldsymbol{\theta}}$ is obtained by only using the first term of this equation
    - The second term is bias $\mathbf{b}$
  - If instead of pre-multiplying with $\boldsymbol{\Psi}^T$, the equation is pre-multiplied with $\mathbf{W}^T$, the following estimate is obtained:
$$\widehat{\boldsymbol{\theta}} = [\mathbf{W}^T\boldsymbol{\Psi}]^{-1}\mathbf{W}^T\mathbf{y}$$
  - Similarly as above, the bias is:
$$\mathbf{b} = [\mathbf{W}^T\boldsymbol{\Psi}]^{-1}\mathbf{W}^T\mathbf{v}$$
- If elements of matrix of instrumental variables $\mathbf{W}$ are uncorrelated with noise $\mathbf{v}$, the bias $\mathbf{b}$ is 0
- Main problem: How to choose elements of matrix $\mathbf{W}$?

Instrumental variables

- Instrumental variables or elements of $\mathbf{W}$ should be:
  - uncorrelated with disturbance, and
  - highly correlated with useful signals $u(k)$ and $y_0(k)$ $\rightarrow$ this requirement is important to guarantee positive definitiveness of $\mathbf{W}^T\boldsymbol{\Psi}$
- Matrix $\boldsymbol{\Psi}$ includes:
  - input signals u that are uncorrelated with noise n, and
  - output signals y that are of course correlated with noise n
- Ideal matrix of instrumental variables $\mathbf{W}$ should include:
  - undisturbed outputs y_0 instead of y, BUT: y_0 cannot be measured $\rightarrow$ it is replaced by instrumental variables that are obtained using the estimated process parameters (usually classical LS estimates $\widehat{\boldsymbol{\theta}}_{LS}$)
- The process model is constructed from the parameters $\widehat{\boldsymbol{\theta}}_{LS}$ $\rightarrow$ its response to the input $u(k)$ is simulated to obtain $y_s(k)$

Instrumental variables

- Algorithm:
  1. Matrix $\boldsymbol{\Psi}$ is constructed from measurements $u(k)$ and $y(k)$
  2. Classical LS parameter estimation gives estimates $\widehat{\boldsymbol{\theta}}_{LS}$:
  $$\widehat{\boldsymbol{\theta}}_{LS} = [\boldsymbol{\Psi}^T\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^T\mathbf{y}$$
  3. System (given by parameters $\widehat{\boldsymbol{\theta}}_{LS}$) is simulated and its response to $u(k)$ is calculated $\rightarrow$ it is denoted with $y_s(k)$
  4. Matrix $\mathbf{W}$ is constructed similarly as $\boldsymbol{\Psi}$ (y_s is used instead of y)
  5. Parameter estimates according to instrumental variables method:
  $$\widehat{\boldsymbol{\theta}}_{IV} = [\mathbf{W}^T\boldsymbol{\Psi}]^{-1}\mathbf{W}^T\mathbf{y}$$
- With non-recursive method the algorithm is iterative (from step 5 we return to step 3 and use parameter estimates $\widehat{\boldsymbol{\theta}}_{IV}$ from step 5 instead of $\widehat{\boldsymbol{\theta}}_{LS}$) – this is repeated until the estimate $\widehat{\boldsymbol{\theta}}_{IV}$ settles (in general small number of iterations)
- Recursive version of the method also exists

Method of stochastic approximation (STA)

- Method of stochastic approximation uses gradient descent algorithm for minimisation of the cost function

$$V(k+1) = \frac{1}{2} e^2(k+1)$$

  - Cost function only takes into account the current equation error:

$$e(k+1) = y(k+1) - \boldsymbol{\psi}^T(k+1)\widehat{\boldsymbol{\theta}}(k)$$

- Gradient descent algorithm gives new parameter estimates in the direction of the steepest descent of the cost function:

$$\widehat{\boldsymbol{\theta}}(k+1) = \widehat{\boldsymbol{\theta}}(k) - \rho(k+1) \frac{d}{d\widehat{\boldsymbol{\theta}}(k)} V(k+1)$$

  - Calculation of the derivative:

$$\frac{d}{d\widehat{\boldsymbol{\theta}}(k)} V(k+1) = e(k+1) \frac{de(k+1)}{d\widehat{\boldsymbol{\theta}}(k)} = -e(k+1)\boldsymbol{\psi}(k+1)$$

Method of stochastic approximation (STA)

- Parameter corrections are calculated according to the formula:

$$\widehat{\boldsymbol{\theta}}(k+1) = \widehat{\boldsymbol{\theta}}(k) + \rho(k+1)e(k+1)\boldsymbol{\psi}(k+1)$$
$$= \widehat{\boldsymbol{\theta}}(k) + \rho(k+1)\boldsymbol{\psi}(k+1)\,[y(k+1) - \boldsymbol{\psi}^T(k+1)\widehat{\boldsymbol{\theta}}(k)]$$

  - The algorithm similar to RLS, but instead of using covariance matrix $\mathbf{P}$ in the update gain, scalar ρ is used
- Method converges if the following requirements for ρ are met:

$$\lim_{k\to\infty}\rho(k) = 0 \qquad \sum_{k=1}^{\infty}\rho(k) = \infty \qquad \sum_{k=1}^{\infty}\rho^2(k) < \infty$$

  - Possible choices for $\rho(k)$ include:
    - $\rho(k) = \frac{1}{k}$
    - $\rho(k) = \mathrm{tr}\{\mathbf{P}(k)\}$ → note that matrix trace is calculated recursively:

$$\rho^{-1}(k+1) = \rho^{-1}(k) + \psi^T(k+1)\psi(k+1), \qquad \rho^{-1}(0) = 0$$

- STA is less computationally demanding than RLS but also provides slower parameter convergence

Parameter estimation – continuous-time processes

- Motivations for parameter estimation of continuous-time models:
  - Parameters of original continuous-time (CT) dynamical system have physical meaning, such as time constant, frequency, damping etc.
  - Transformation: CT dynamical system $\rightarrow$ discrete-time (DT) domain:
    - The parameters of DT systems are nonlinear mapping of CT parameters
    - Their interpretation is much more difficult $\rightarrow$ harder validation
    - Especially in the case when some CT parameters are known and only a few are unknown, the use of identification of DT models not useful
  - No need to choose sampling time T – the parameters T-independent
  - The sampling of signals not required to be equidistant
- It is possible to find CT equivalents of DT models, but these nonlinear mappings can cause some unpredictable problems
- The idea of CT parameter estim. – the same as in the case of DT:
  - reformulating the problem as overdetermined system of equations

Parameter estimation – continuous-time processes

- Continuous-time transfer function of the process:

$$G_p(s) = \frac{Y(s)}{U(s)} = \frac{b_0 + b_1 s + \cdots + b_n s^n}{1 + a_1 s + \cdots + a_n s^n} \qquad a_n \neq 0$$

- The system described in the form of a differential equation:

$$\sum_{i=0}^{n} a_i \frac{d^i}{dt^i} y(t) = \sum_{i=0}^{n} b_i \frac{d^i}{dt^i} u(t) \qquad a_0 = 1$$

- Expressing $y(t)$ from the above differential equation :

$$y(t) = -a_1 \frac{d}{dt} y(t) \ldots - a_n \frac{d^n}{dt^n} y(t) + b_0 u(t) + b_1 \frac{d}{dt} u(t) + \cdots + b_n \frac{d^n}{dt^n} u(t)$$

- Right-hand side can be written as a product between vector of regressors (holding measured signals and their derivatives) and parameter vector
  - BUT: The derivatives are not measurable; their approximations amplify noise which leads to identification results deterioration

Parameter estimation – continuous-time processes

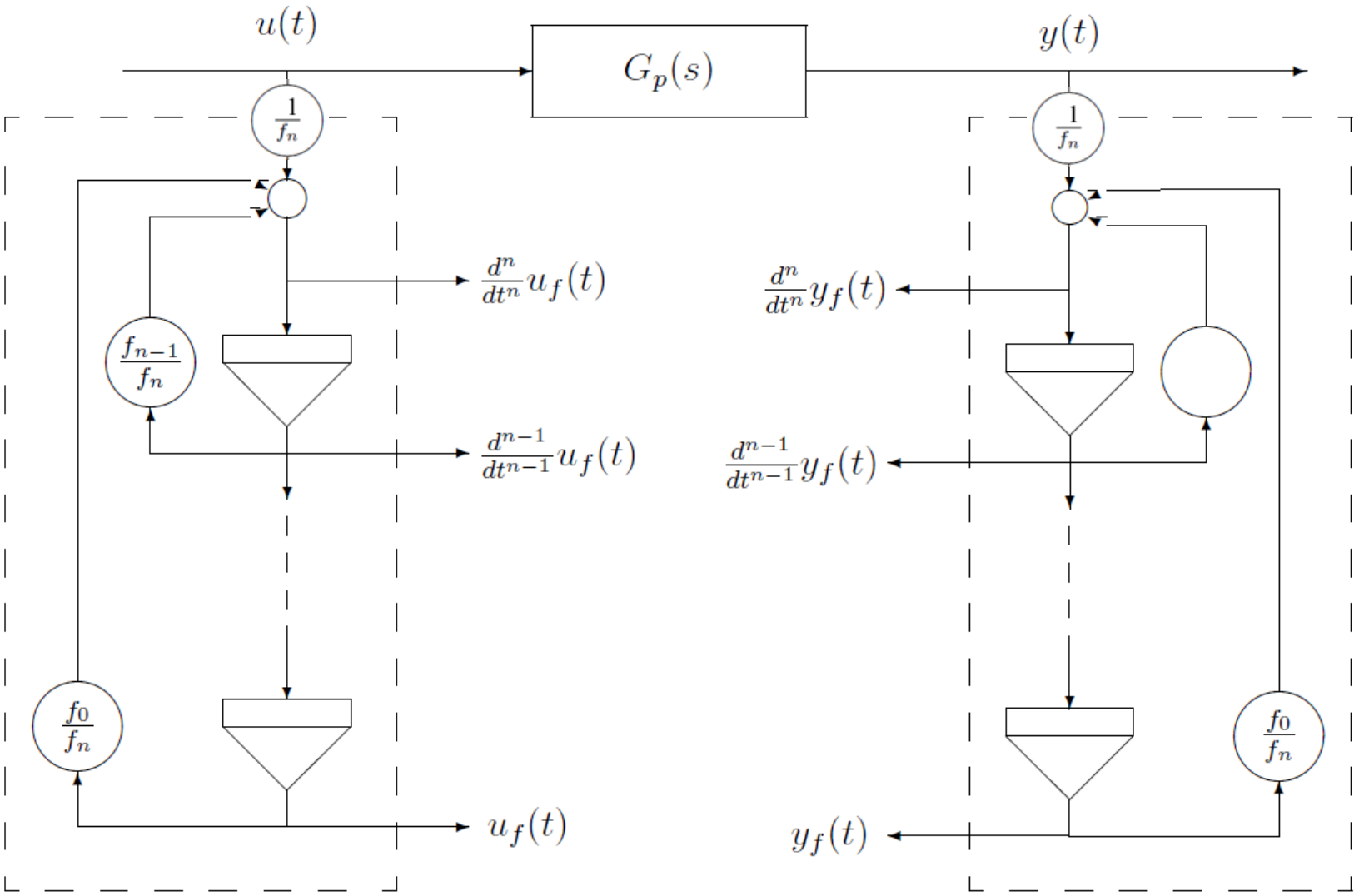
- An alternative solution is to introduce state variable filters:
  - low-pass filter that dampens out the high-frequency noise on u and y
  - transforms the system into a state-space representation
- The idea: $u(t) \xrightarrow{G_f(s)} u_f(t)$ and $y(t) \xrightarrow{G_f(s)} y_f(t)$
  - The transfer function between original signals is the same as the transfer function between the filtered signals for any filter $G_f(s)$
- The difference between orders of denominator and numerator of $G_f(s) \rightarrow$ must be higher (or equal) than the process order n
- Simplest choice – denominator order is n, numerator order is 0:

$$G_f(s) = \frac{1}{f_0 + f_1 s + \cdots + f_n s^n} \qquad f_n \neq 0$$

- Laplace transform of filtered input and output:

$$U_f(s) = G_f(s)U(s) \qquad Y_f(s) = G_f(s)Y(s)$$

Parameter estimation – continuous-time processes



Parameter estimation – continuous-time processes

- The relation between filtered signals is the same as the relation between original signals:

$$\frac{Y_f(s)}{U_f(s)} = \frac{G_f(s)Y(s)}{G_f(s)U(s)} = \frac{Y(s)}{U(s)} = G_p(s)$$

- Description in the form of a differential equation:

$$y_f(t) = -a_1 \frac{d}{dt} y_f(t) - \cdots - a_n \frac{d^n}{dt^n} y_f(t) + b_0 u_f(t) + b_1 \frac{d}{dt} u_f(t) + \cdots + b_n \frac{d^n}{dt^n} u_f(t)$$

  - Differential equation between filtered signals has the same structure and the same parameters as the original differential equation

- A new vector of estimated parameter is introduced together with the corresponding vector of regressors:

$$\boldsymbol{\theta}^T = [a_1 \quad \dots \quad a_n \quad b_0 \quad b_1 \quad \dots \quad b_n]$$

$$\boldsymbol{\psi}^T(t) = \left[-\frac{d}{dt} y_f(t) \quad \dots \quad -\frac{d^n}{dt^n} y_f(t) \quad u_f(t) \quad \frac{d}{dt} u_f(t) \quad \dots \quad \frac{d^n}{dt^n} u_f(t)\right]$$

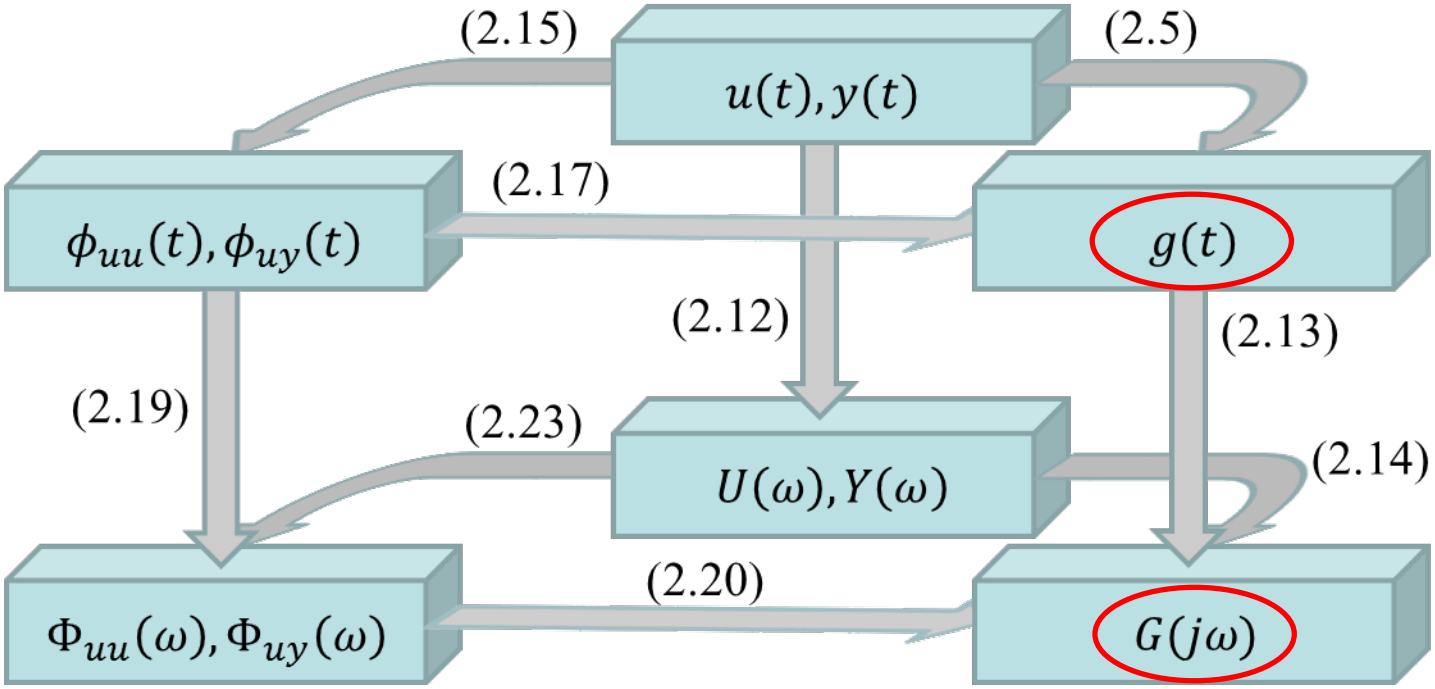
$$y_f(t) = \boldsymbol{\psi}^T(t)\boldsymbol{\theta}$$

Parameter estimation – continuous-time processes

- What follows is a standard parameter estimation procedure
- Non-recursive method:
  - From N observations at times $t_1, t_2, \ldots t_N$ we construct:
    - matrix $\boldsymbol{\Psi}$ – in its i-th row there is row vector $\boldsymbol{\psi}^T(t_i)$
    - vector $\mathbf{y}$ – in its i-th row there is scalar $y_f(t_i)$
  - The solution is a standard LS parameter estimate: $\widehat{\boldsymbol{\theta}} = [\boldsymbol{\Psi}^T\boldsymbol{\Psi}]^{-1}\boldsymbol{\Psi}^T\mathbf{y}$
- Recursive version also exists
- Note that $G_f(s)$ is a continuous filter – analog implementation
- Alternatively, $G_f(s)$ can be implemented by means of simulation on a digital computer, but some care is needed with I/O signals:
  - Input signal – usually constant between two samples (D/A converter)
  - Output signal – usually changes between two samples (in case of fast sampling linear interpolation can be assumed)

Identification of non-parametric models

- Identification of non-parametric models is based on this scheme:



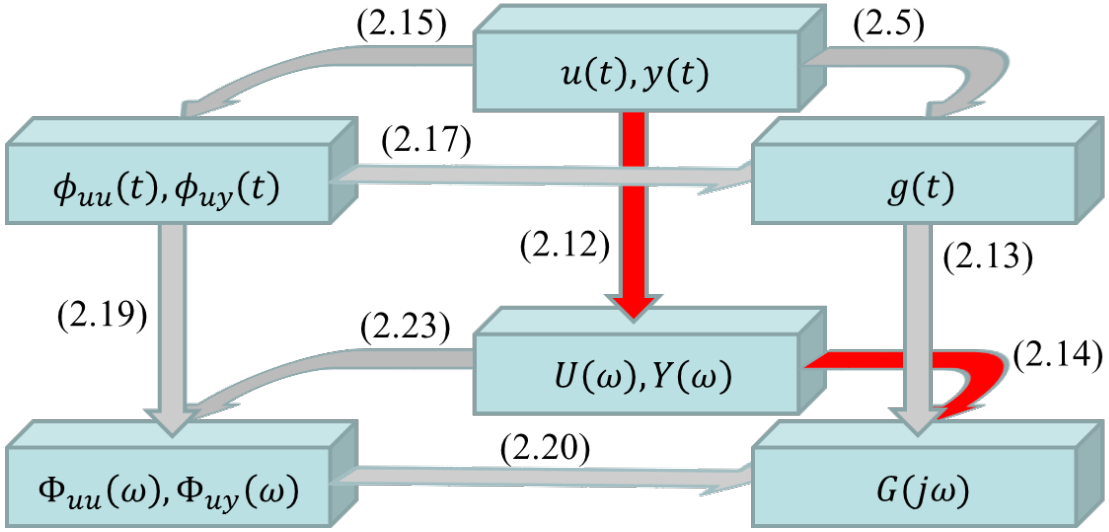
- Relation between process signals, their Fourier transforms, signal correlations, signal spectra, and non-parametric models is shown – the identification of the latter will be treated in the following

Identification of non-parametric models

- Each method uses a characteristic path from process signals to a non-parametric model in the figure
- The following methods will be dealt with:
  - Empirical transfer function estimation (ETFE) or Fourier analysis
  - Analysis of frequency response (direct and correlation-function based)
  - Correlation analysis
  - Spectral analysis
- The differences among the above methods:
  - The class of input signals that are used for the excitation:
    - aperiodic
    - sinusoidal or more generally periodic
    - random or pseudo-random
  - The level of sensitivity to disturbances and noise

Empirical transfer function estimation (ETFE)

- The most direct and obvious non-parametric identification method is calculation of frequency response based on division of Fourier transforms of input and output signals



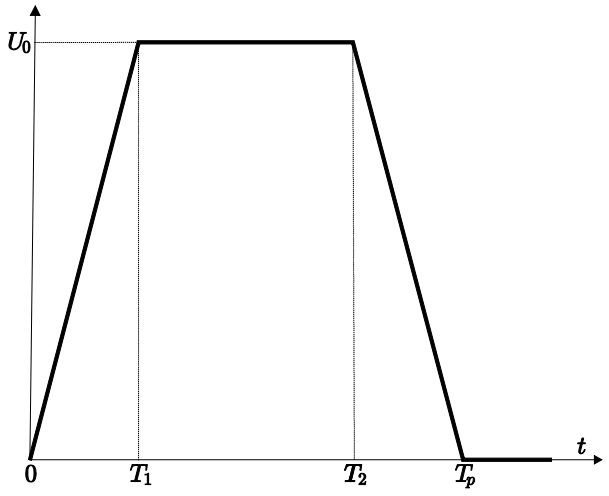
- Most often aperiodic input signals are used – usually a series of pulses of different length (a step is a long pulse)

ETFE – Fourier transform of simple signals

- Example: Calculate Fourier transform of:
  - rectangular pulse
  - triangular pulse
  - trapezoidal pulse
  - point-symmetric rectangular pulse
- What is the influence of pulse width to the Fourier transform?

  - Trapezoidal pulse (T_p is the length, T_1 is rise time, T_2 the time when it starts to descent):

$$U_{tra}(j\omega) = U_0 T_2 \left[\frac{\sin\frac{\omega T_1}{2}}{\frac{\omega T_1}{2}}\right]\left[\frac{\sin\frac{\omega T_2}{2}}{\frac{\omega T_2}{2}}\right] e^{-\frac{j\omega T_p}{2}}$$



    - Spectrum zeros: $\frac{\omega_{1n} T_1}{2} = n\pi$ or altern. $f_{1n} = \frac{n}{T_1}$

      $\frac{\omega_{2n} T_2}{2} = n\pi$ or altern. $f_{2n} = \frac{n}{T_2}$ $\quad n = 1,2,3 \ldots$

ETFE – Fourier transform of simple signals

- Rectangular pulse (T_p is pulse width):
  - Rectangular pulse is obtained from trapezoidal pulse by setting $T_1 = 0$ and $T_2 = T_p$:

$$U_{rec}(j\omega) = U_0 T_p \left[\frac{\sin \frac{\omega T_p}{2}}{\frac{\omega T_p}{2}} \right] e^{-\frac{j\omega T_p}{2}}$$

The first zero of the spectrum: $\frac{\omega_z T_p}{2} = \pi$ or $f_z = \frac{\omega}{2\pi} = \frac{2\pi}{2\pi T_p} = \frac{1}{T_p}$

- Triangular pulse (T_p is pulse width, the pulse is symmetric):
  - Obtained from trapezoidal pulse by setting $T_1 = T_p/2$ and $T_2 = T_p/2$:

$$U_{tri}(j\omega) = U_0 \frac{T_p}{2} \left[\frac{\sin \frac{\omega T_p}{4}}{\frac{\omega T_p}{4}} \right]^2 e^{-\frac{j\omega T_p}{2}}$$

The first zero of the spectrum (double): $\frac{\omega_z T_p}{4} = \pi$ or $f_z = \frac{\omega}{2\pi} = \frac{2}{T_p}$

ETFE – Fourier transform of simple signals

- Point-symmetric rectangular pulse:

$$U_{rec\_sym}(j\omega) = \frac{U_0}{2} T_p \left[\frac{\sin^2 \frac{\omega T_p}{4}}{\frac{\omega T_p}{4}} \right] e^{-j\frac{\omega T_p - \pi}{2}}$$

  - Because it has zero mean, it does not excite very low frequencies
  - Besides spectrum zero at $\omega = 0$, it also has all the zeros of the triangular signal

- This task can also be performed in MATLAB 

ETFE – identification procedure

- Identification procedure for empirical TF estimation:
  - An appropriate range of excitation frequencies is estimated:
    - This always includes natural frequencies of the process (defined by dominant time constants)
    - The estimated frequency-response error will be large around frequencies with no or low excitation
    - The excitation of a too large frequency band is also not good because the over-all power is usually limited and not enough power is left for „important“ frequencies
    - The length of the measurement also has to be estimated – this defines frequency resolution when using FFT
  - Excitation is applied to the process – the output is measured
  - Fourier transforms of input and output, respectively, are needed:

$$Y(\omega) = \int_{-\infty}^{\infty} y(t) e^{-j\omega t} dt \qquad U(\omega) = \int_{-\infty}^{\infty} u(t) e^{-j\omega t} dt$$

ETFE – identification procedure

- The result of the identification, i.e., the frequency response, is obtained by dividing the Fourier transforms:

$$G(j\omega) = \frac{Y(\omega)}{U(\omega)}$$

- Verification
- Validation
- If necessary certain steps of the algorithm are repeated

- MATLAB examples:
  - Long pulse → problems of zeros in the spectrum
  - Frequency resolution not adequate
  - Impulse response not smooth, errors in frequency response
  -

ETFE – convergence analysis – effect of noise

- Output signal $y(t)$ is corrupted with some aditive disturbance $n(t)$:

$$y(t) = y_0(t) + n(t)$$

- The resulting frequency response:

$$G(j\omega) = \frac{\mathcal{F}\{y_0(t) + n(t)\}}{\mathcal{F}\{u(t)\}} = G_0(j\omega) + \Delta G_n(j\omega)$$

- Frequency-response error:

$$\Delta G_n(j\omega) = \frac{\mathcal{F}\{n(t)\}}{\mathcal{F}\{u(t)\}}$$

$$\Delta G_n(j\omega) = \frac{N(\omega)}{U(\omega)}$$

ETFE – convergence analysis – effect of noise

Bias

- If $n(t)$ is not correlated with input $u(t)$ and undisturbed output signal $y_0(t)$, and its mean value is 0 ($E\{n(t)\} = 0$), the mathematical expectation of frequency-response error $\Delta G_n(j\omega)$ becomes:

$$E\{\Delta G_n(j\omega)\} = E\left\{\frac{\mathcal{F}\{n(t)\}}{\mathcal{F}\{u(t)\}}\right\} = \frac{E\{\mathcal{F}\{n(t)\}\}}{E\{\mathcal{F}\{u(t)\}\}} = \frac{\mathcal{F}\{E\{n(t)\}\}}{\mathcal{F}\{E\{u(t)\}\}} = \frac{0}{U(\omega)}$$

- The estimated frequency response is **unbiased** for all excited frequencies, i.e., where $U(\omega) \neq 0$
- The estimate of frequency response is also consistent because it is bias-free

ETFE – convergence analysis – effect of noise

Consistency in the mean square

- Variance of the frequency-response error $\Delta G_n(j\omega)$:

$$E\{|\Delta G_n(j\omega)|^2\} = \frac{E\{\mathcal{F}\{n(t)\}\overline{\mathcal{F}\{n(t)\}}\}}{E\{\mathcal{F}\{u(t)\}\overline{\mathcal{F}\{u(t)\}}\}} = \frac{E\{\int_{-\infty}^{\infty} n(t)e^{-j\omega t}dt \int_{-\infty}^{\infty} n(\tau)e^{+j\omega\tau}d\tau\}}{E\{\int_{-\infty}^{\infty} u(t)e^{-j\omega t}dt \int_{-\infty}^{\infty} u(\tau)e^{+j\omega\tau}d\tau\}} =$$

$$= \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\{n(t)n(\tau)\}e^{-j\omega(t-\tau)}dtd\tau}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\{u(t)u(\tau)\}e^{-j\omega(t-\tau)}dtd\tau} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_{nn}(t-\tau)e^{-j\omega(t-\tau)}dtd\tau}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_{uu}(t-\tau)e^{-j\omega(t-\tau)}dtd\tau} =$$

$$= \frac{\Phi_{nn}(\omega)}{\Phi_{uu}(\omega)} \approx \frac{\Phi_{nn}(\omega)T_A}{|U(\omega)|^2}$$

  - The overbar in the above equation denotes complex conjugate of a complex number
  - T_A is the observation time that has to be long enough to include all the transients

ETFE – convergence analysis – effect of noise

$$E\{|\Delta G_n(j\omega)|^2\} \approx \frac{\Phi_{nn}(\omega) T_A}{|U(\omega)|^2}$$

- Variance of the frequency-response error at certain frequency ω:
  - increases with increased power spectral density of the noise at ω
  - increases with longer observation interval T_A
  - decreases with higher power spectral density of the input at ω
- The second bullet above might not seem logical at first:
  - one would expect better results if observation time is longer
  - but longer observation does not bring any new information about the transient (or system dynamics) – useful signal energy does not increase
  - on the other hand longer observation does increase the energy of the noise
- Note that the variance of the unexcited frequencies ($|U(j\omega)| \to 0$) is theoretically unbounded

ETFE – convergence analysis – effect of noise

- Noise is white → the variance of the frequency-response error:

$$E\{|\Delta G_v(j\omega)|^2\} = \frac{\Phi_0 T_A}{|U(\omega)|^2}$$

  - The estimated frequency response is un-biased
  - But in example 5 it looks biased
  - The estimated frequency response is low in amplitude at high frequencies and the noise component becomes dominant
  - Amplitude response is obtained by taking absolute value of frequency response (it is shown in logarithmic scale)
  - Mean-value of the absolute value of noise is not zero (!):
    - It can be shown that mean-value of the absolute value of noise $v(t)$ with noise distribution $p(v)$ is $\int_{-\infty}^{\infty} |v| p(v) dv$
    - For Gaussian noise with standard deviation σ_v the mean value of $|v(t)|$ is $\sigma_v \sqrt{\frac{2}{\pi}} = 0.7979\sigma_v$

ETFE – convergence analysis – unknown op. point

- Signals $u(t)$ and $y(t)$ are only deviations of the measured or absolute signals $U(t)$ and $Y(t)$ from their respective stationary values U_{00} and Y_{00} (that define the operating point):

$$u(t) = U(t) - U_{00}$$
$$y(t) = Y(t) - Y_{00}$$

- Stationary value Y_{00} of the output signal is estimated before the start or after the end of the „identification signal“ (these two values can be different in case of the process with integral character)
- The estimate of Y_{00} can be obtained by averaging absolute output signal $Y(t)$ before the start or after the end of the transient
  - The case „before“ (the length of the averaging interval is T_B):

$$\hat{Y}_{00} = \frac{1}{T_B}\int_{-T_B}^{0} Y(t)dt$$

ETFE – convergence analysis – unknown op. point

- Mathematical expectation of the error of the stationary state estimate $\Delta Y_{00} = \hat{Y}_{00} - Y_{00}$ is ($y_0(t)$ is 0 in the operating point):

$$E\{\Delta Y_{00}\} = E\left\{\frac{1}{T_B}\int_{-T_B}^{0}\left[Y_{00} + y_0(t) + n(t)\right]dt - Y_{00}\right\} = \frac{1}{T_B}\int_{-T_B}^{0} E\{n(t)\}dt = 0$$

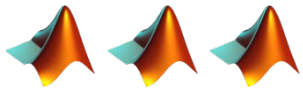
  - The estimate is **bias-free** if mean value of the noise is 0
- The variance of the stationary-state-estimate error:

$$E\{\Delta Y_{00}^2\} = E\left\{\left[\frac{1}{T_B}\int_{-T_B}^{0} n(t)dt\right]^2\right\} = \frac{1}{T_B^2}\int_{-T_B}^{0}\int_{-T_B}^{0} E\{n(t)n(\tau)\}dtd\tau = \frac{1}{T_B^2}\int_{-T_B}^{0}\int_{-T_B}^{0} \phi_{nn}(t-\tau)dtd\tau$$

  - For white noise ($n(t) = v(t)$) the above expression reduces to:

$$E\{\Delta Y_{00}^2\} = \frac{\Phi_0}{T_B}$$

- The error due to unknown operating point:
  - Step with amplitude $\Delta Y_{00} \rightarrow$ its Fourier transform is $\frac{\Delta Y_{00}}{j\omega}$
  - The error of the frequency response is $\Delta G_{\Delta Y}(j\omega) = \frac{\Delta Y_{00}}{j\omega U(\omega)}$



ETFE – convergence analysis

- Variance of the stationary-state-estimate error for white noise is therefore:

$$E\{|\Delta G_{\Delta Y}(j\omega)|^2\} = \frac{\Phi_0}{\omega^2 |U(\omega)|^2 T_B}$$

- If ETFE is repeated m times and the results of the individual estimates are averaged, the effect is the same as with m times higher power spectral density of the excitation:

$$m \frac{|U(\omega)|^2}{T_A}$$

- Both sources of error (white noise at the process output and unknown operating point) are combined $\rightarrow$ the variance of the frequency-response-estimate error:

$$E\{|\Delta G(j\omega)|^2\} = \frac{\Phi_0}{m|U(\omega)|^2}\left(T_A + \frac{1}{\omega^2 T_B}\right)$$

ETFE – the choice of appropriate input signals

- The appropriate input signals should satisfy:
  - Being implementable (depends heavily on the equipment used)
  - Power spectral density should be as high as possible in the frequency interval defined by the designer BUT respecting the bounded amplitude of the signal
- Input signal should be chosen such that it minimises the estimated frequency response error in some sense
- The signals that satisfy these requirements:
  - At low frequencies – step signals (in practical implementations long pulses are used)
  - At medium and high frequencies rectangular pulses are used
  - PRBS – it has constant PSD (except at zero frequency)
  - Chirp or sweep signal (where frequency changes with time) – implementation is more difficult – continuous amplitude distribution

ETFE – the choice of appropriate input signals

- The required amplitude density can be defined:

$$|U(\omega)|_{req} = \frac{\sqrt{\Phi_{nn} T_A}}{\sigma_G(\omega)\sqrt{m}}$$

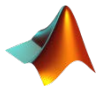
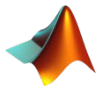
  - m is the number of experiment repetitions
  - $\sigma_G(\omega)$ is the allowable error of the frequency response estimate
- The above equation shows that the required amplitude density depends heavily on the level of noise and on the model purpose (that is quantified by $\sigma_G(\omega)$):
  - General guidelines for the required amplitude density are therefore not possible without the detailed knowledge of the process
  - For the controller design it is most important to have the error in frequency response estimate low at medium frequencies where phase and gain margin of the control loop are determined

ETFE – the choice of appropriate input signals

- Nevetheless, very often a series of rectangular pulses is used:
  - Longer pulses for determination of low frequencies – 20-30% of time
  - Shorter pulses .....medium..... – 70-80% of time
  - If the frequency response is to be estimated on the interval $[0, \omega_{max}]$, the length of the **shortest** pulse can be determined as:
    - Rectangular pulse has adequate excitation until one half of the frequency of the first zero in the signal spectrum ($\omega_z = 2\pi/T_p$)
    - The length of the pulse T_p is therefore selected so that:

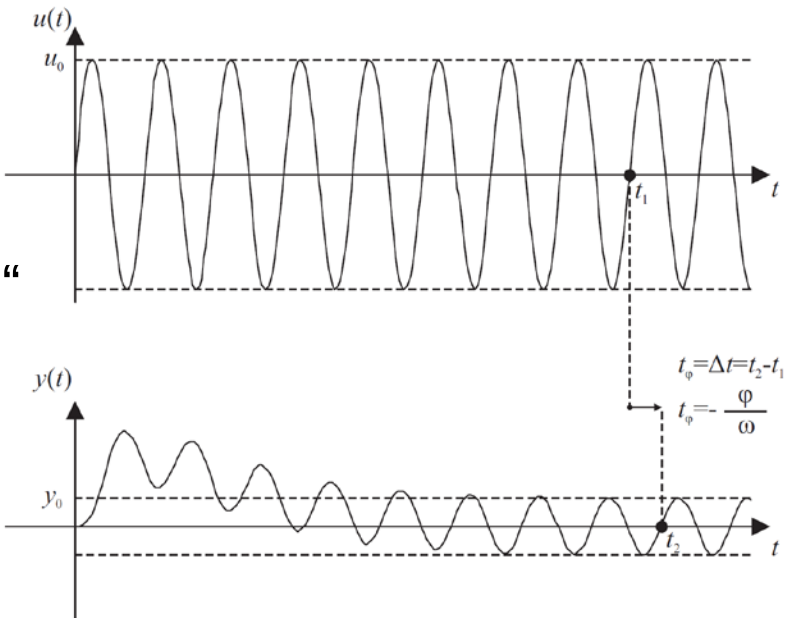
$$\omega_{max} = \frac{\omega_z}{2} = \frac{\pi}{T_p} \quad \Rightarrow \quad T_p = \frac{\pi}{\omega_{max}}$$

  - The input signals composed of series of rectangular pulses are similar to PRBS

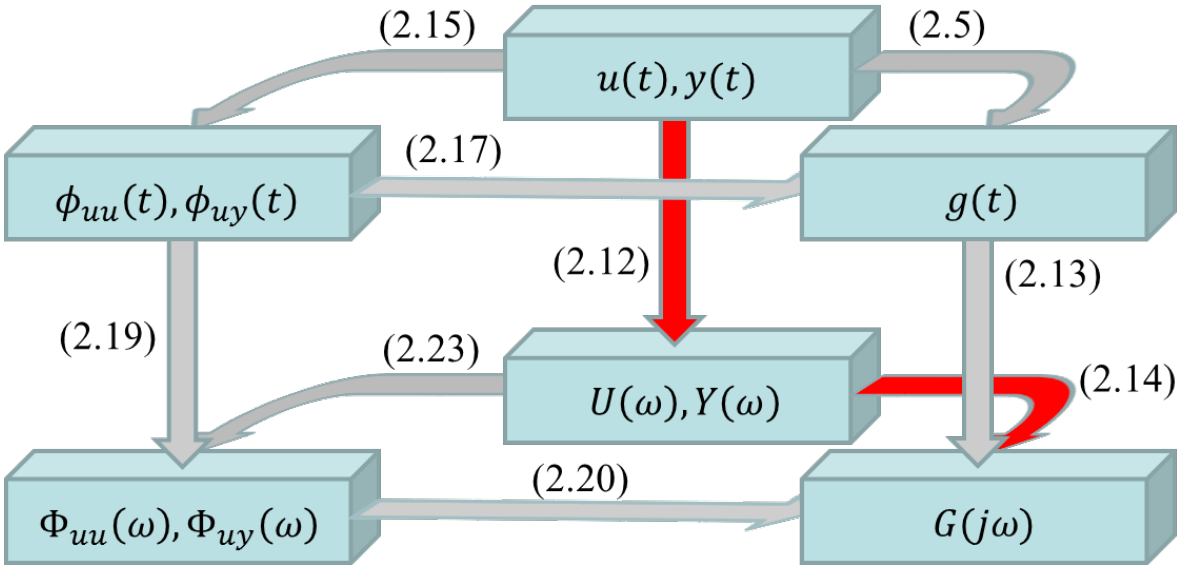
Analysis of frequency response

- Frequency response $G(j\omega)$ is a complex function of frequency:
  - Its absolute value $|G(j\omega)|$ is called amplitude response – it defines the output/input amplitude quotient when excitation is harmonic
  - Its phase angle $\sphericalangle G(j\omega)$ is called phase response – it defines phase shift of output oscillation w.r.t input that is harmonic
- Frequency response can be measured according to its definition:
  - The input is excited by a sine/cosine of some frequency ω_0
  - When transient decays, amplitude quotient and phase difference are measured
  - The experiment has to be repeated at more frequencies in the „interesting“ frequency interval to obtain „sampled“ estimate of the frequency response



Analysis of frequency response

- This identification method takes the same path as ETFE in the figure below but uses periodic excitation instead of aperiodic



- Because only one frequency is present in the signal, calculation of Fourier transform in the classical sense is not necessary

Analysis of frequency response

- More approaches to measurement of amplitude/phase:
  - Classical approach with oscilloscope
  - Two-channel plotter
  - Special device – compensator – it has two sinusoidal oscillations at the output – a test one and a comparative one
  - All these approaches suitable for the cases with low noise
- An alternative approach is to estimate the frequency response at ω_0 by using a **sampler** that takes only two samples of the output:
  - The input is a cosine signal (it could also be a sine):

$$u(t) = U_0 \cos(\omega_0 t)$$

  - Again we have to wait until the end of the transient when the output oscillation is harmonic:

$$y(t) = U_0 |G(j\omega_0)| \cos[\omega_0 t + \varphi(\omega_0)] \qquad t > t_{transient}$$

Analysis of frequency response with a sampler

- The output signal is sampled in two sampling instants:
  - The first one (t_1) is aligned with a start of a new period (of the input):
$$t_1 = nt_p = n\frac{2\pi}{\omega_0} \quad n = 1,2,3,\ldots$$
  - The second one (t_2) is taken a quarter of the period before:
$$t_2 = nt_p - \frac{t_p}{4} = \left(n - \tfrac{1}{4}\right)\frac{2\pi}{\omega_0} \quad n = 1,2,3,\ldots$$
  - $t_p = \frac{2\pi}{\omega_0}$ is a period of harmonic oscillations with angular frequency ω_0
- The following mathematical relations are used:
$$\cos(\omega_0 t_1 + \varphi) = \cos(2\pi n + \varphi) = \cos\varphi$$
$$\cos(\omega_0 t_2 + \varphi) = \cos(2\pi n - \tfrac{\pi}{2} + \varphi) = \sin\varphi$$
- Output samples at t_1 and t_2 are therefore:
$$y(t_1) = U_0|G(j\omega_0)|\cos\varphi(\omega_0)$$
$$y(t_2) = U_0|G(j\omega_0)|\sin\varphi(\omega_0)$$

Analysis of frequency response with a sampler

- Complex frequency response can be decomposed into its real and imaginary parts that are related to its amplitude and phase:

$$\Re[(Gj\omega_0)] = |G(j\omega_0)| \cos\varphi(\omega_0)$$
$$\Im[G(j\omega_0)] = |G(j\omega_0)| \sin\varphi(\omega_0)$$

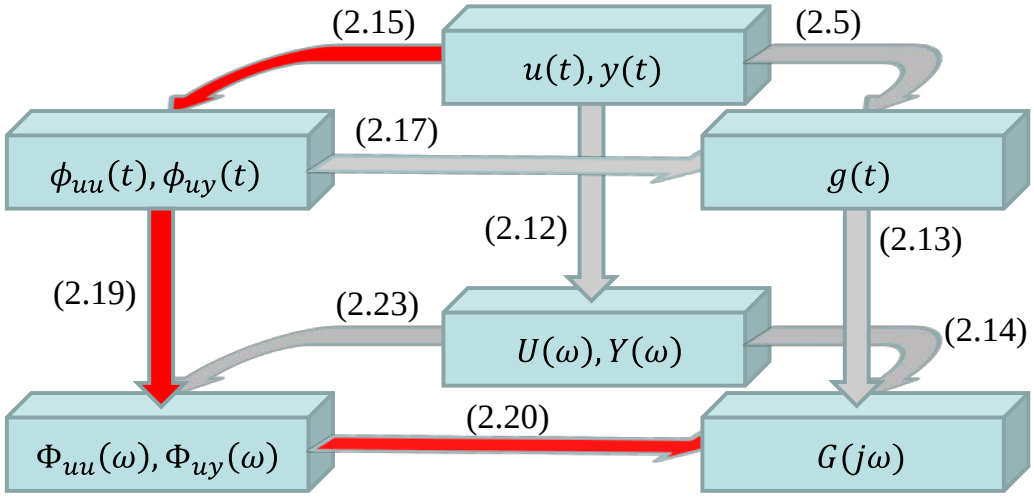
- Real and imaginary part can be obtained from output samples:

$$y(t_1) = U_0\Re[(Gj\omega_0)] \quad \rightarrow \quad \Re[(Gj\omega_0)]$$
$$y(t_2) = U_0\Im[G(j\omega_0)] \quad \rightarrow \quad \Im[(Gj\omega_0)]$$

- If noise is zero-mean, frequency response estimate is **bias-free**
- The variances of real and imaginary parts are directly related to the variance of the output measurements (only divided with U_0^2):
  - Only useful for processes with low level of noise
  - Only two measurements at each frequency
    $\rightarrow$ consistency in the mean square
    cannot be defined

Analysis of frequency response

- If signals are (very) noisy, correlation methods come useful:
  - Orthogonal correlation is an approach that can be classified into the group of methods for analysis of frequency response
- Orthogonal correlation takes the red path in the scheme:
  - Again sine/cosine input signals are used
  - Again one experiment is needed for estimation of one point of frequency response → Fourier analysis is not needed



Orthogonal correlation

- An arbitrary phase of the input signal can be chosen – here a cosine is assumed to make the following equations simpler:

$$u(t) = U_0 \cos(\omega_0 t)$$

- The stationary part of the output is:

$$y(t) = U_0 |G(j\omega_0)| \cos[\omega_0 t + \varphi(\omega_0)] \qquad t > t_{transient}$$

- Auto-correlation of u and cross-correlation between u and y are:

$$\phi_{uu}(\tau) = E\{u(t)u(t+\tau)\} = \lim_{T_i \to \infty} \frac{1}{T_i} \int_0^{T_i} u(t)u(t+\tau)dt$$

$$\phi_{uy}(\tau) = E\{u(t)y(t+\tau)\} = \lim_{T_i \to \infty} \frac{1}{T_i} \int_0^{T_i} u(t)y(t+\tau)dt$$

Orthogonal correlation

- The estimates of auto-correlation and cross-correlation are obtained by integrating periodic functions over integer number of periods:

$$\hat{\phi}_{uu}(\tau) = \frac{1}{nt_p}\int_0^{nt_p} u(t)u(t+\tau)dt =$$

$$= \frac{1}{nt_p}\int_0^{nt_p} U_0\cos(\omega_0 t)U_0\cos[\omega_0(t+\tau)]dt = \frac{U_0^2}{2}\cos(\omega_0\tau)$$

$$\hat{\phi}_{uy}(\tau) = \frac{1}{nt_p}\int_0^{nt_p} u(t)y(t+\tau)dt =$$

$$= \frac{1}{nt_p}\int_0^{nt_p} U_0\cos(\omega_0 t)U_0|G(j\omega_0)|\cos[\omega_0(t+\tau)+\varphi(\omega_0)]dt$$

$$= \tfrac{U_0^2}{2}|G(j\omega_0)|\cos[\omega_0\tau+\varphi(\omega_0)]$$

Orthogonal correlation

- Auto-correlation and cross-correlation are evaluated for two distinct values of its argument τ ($\tau = 0$ in $\tau = -\frac{t_p}{4} = -\frac{\pi}{2\omega_0}$):

$$\hat{\phi}_{uy}(0) = \frac{U_0^2}{2}|G(j\omega_0)|\cos[\varphi(\omega_0)] = \frac{U_0^2}{2}\Re[G(j\omega_0)]$$

$$\hat{\phi}_{uy}\left(\frac{-\pi}{2\omega_0}\right) = \frac{U_0^2}{2}|G(j\omega_0)|\cos\left[-\frac{\pi}{2} + \varphi(\omega_0)\right] =$$
$$= \frac{U_0^2}{2}|G(j\omega_0)|\sin[\varphi(\omega_0)] = \frac{U_0^2}{2}\Im[G(j\omega_0)]$$

- From the estimates $\hat{\phi}_{uy}(0)$ and $\hat{\phi}_{uy}\left(\frac{-\pi}{2\omega_0}\right)$, real and imaginary parts of the frequency response at ω_0 can be obtained easily

Orthogonal correlation

- Real and imaginary part of frequency response are obtained by taking into account equations for correlation functions:

$$\Re[G(j\omega_0)] = \frac{2}{U_0^2 n t_p}\int_0^{nt_p} U_0\cos(\omega_0 t)y(t)dt = \frac{2}{U_0 n t_p}\int_0^{nt_p} y(t)\cos(\omega_0 t)dt$$

$$\Im[G(j\omega_0)] = \frac{2}{U_0^2 n t_p}\int_0^{nt_p} U_0\cos(\omega_0 t)y\left(t-\frac{\pi}{2\omega_0}\right)dt =$$

$$= \frac{2}{U_0 n t_p}\int_{-\frac{\pi}{2\omega_0}}^{nt_p-\frac{\pi}{2\omega_0}} y(t)[-\sin(\omega_0 t)]dt$$

- Due to periodicity the integral bounds can be shifted to 0 and nt_p in the last equation:

$$\Im[G(j\omega_0)] = \frac{-2}{U_0 n t_p}\int_0^{nt_p} y(t)\sin(\omega_0 t)dt$$

Orthogonal correlation – convergence analysis

- Some disturbance $n(t)$ is always present on the output signal:

$$y(t) = y_0(t) + n(t)$$

- This leads to the error in the frequency response estimation:

$$\Re[G(j\omega_0)] = \frac{2}{U_0 n t_p}\int_0^{nt_p}[y_0(t)+n(t)]\sin(\omega_0 t)dt = \Re[G_0(j\omega_0)] + \Delta\Re[G(j\omega_0)$$

$$\Im[G(j\omega_0)] = \frac{2}{U_0 n t_p}\int_0^{nt_p}[y_0(t)+n(t)]\cos(\omega_0 t)dt = \Im[G_0(j\omega_0)] + \Delta\Im[G(j\omega_0)]$$

- Due to linearity, the error equations are similar to the estimates:

$$\Delta\Re[G(j\omega_0)] = \frac{2}{U_0 n t_p}\int_0^{nt_p} n(t)\cos(\omega_0 t)dt$$

$$\Delta\Im[G(j\omega_0)] = \frac{-2}{U_0 n t_p}\int_0^{nt_p} n(t)\sin(\omega_0 t)dt$$

Orthogonal correlation – convergence analysis

- If noise $n(t)$ is zero-mean ($E\{n(t)\} = 0$) and it is not correlated with the input, the error of the estimated frequency response becomes:

$$E\{\Delta\Re[G(j\omega_0)]\} = E\left\{\frac{2}{U_0 n t_p}\int_0^{nt_p} n(t)\cos(\omega_0 t)\,dt\right\} =$$

$$= \frac{2}{U_0 n t_p}\int_0^{nt_p} E\{n(t)\}\cos(\omega_0 t)dt = 0$$

$$E\{\Delta\Im[G(j\omega_0)]\} = E\left\{\frac{-2}{U_0 n t_p}\int_0^{nt_p} n(t)\sin(\omega_0 t)\,dt\right\} =$$

$$= \frac{-2}{U_0 n t_p}\int_0^{nt_p} E\{n(t)\}\sin(\omega_0 t)dt = 0$$

- The estimate of the frequency response is then **bias-free**

Orthogonal correlation – convergence analysis

- Variance of the frequency response error is:

$$E\{|\Delta G(j\omega_0)|^2\} = E\{\Delta\Re^2[G(j\omega_0)] + \Delta\Im^2[G(j\omega_0)]\}$$

- Three types of disturbance signals will be treated:
  - High-frequency quasi-stationary noise
  - Harmonic disturbances (sine, cosine)
  - Low-frequency non-stationary disturbances (drift)

- The variances for these three types of disturbances can be obtained analytically after long derivations if certain assumptions are taken into account

- In the following only the resulting variances will be given

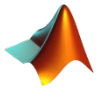
Orthogonal correlation – convergence analysis

- High-frequency quasi-stationary noise:

$$\sigma_{G_n}^2(\omega_0) = E\{|\Delta G(j\omega_0)|^2\} = \frac{4}{U_0^2 n T_p}\int_{-nT_p}^{nT_p} \phi_{nn}(\tau)\left[1-\frac{|\tau|}{nT_p}\right]e^{-j\omega_0\tau}d\tau$$

- For white noise ($n(t) = v(t)$) with power spectral density Φ_0 the following variance of the frequency response error is obtained:

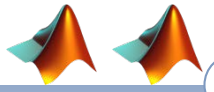
$$\sigma_{G_v}^2 = \frac{4\Phi_0}{U_0^2 n T_p}$$



- For broad-band noise and long measurement times:

$$\sigma_{G_n}^2(\omega_0) \approx \frac{4}{U_0^2 n T_p}\int_{-nT_p}^{nT_p} \phi_{nn}(\tau)e^{-j\omega_0\tau}d\tau \approx \frac{4\Phi_{nn}(\omega_0)}{U_0^2 n T_p}$$

- Method is consistent in the mean square:
  - Variance of the estimate error converges to 0 as $n \to \infty$



Orthogonal correlation – convergence analysis

- The relation between the PSD of the signal at the noise filter output $\Phi_{nn}(\omega)$ and the one at the filter input Φ_0 is:

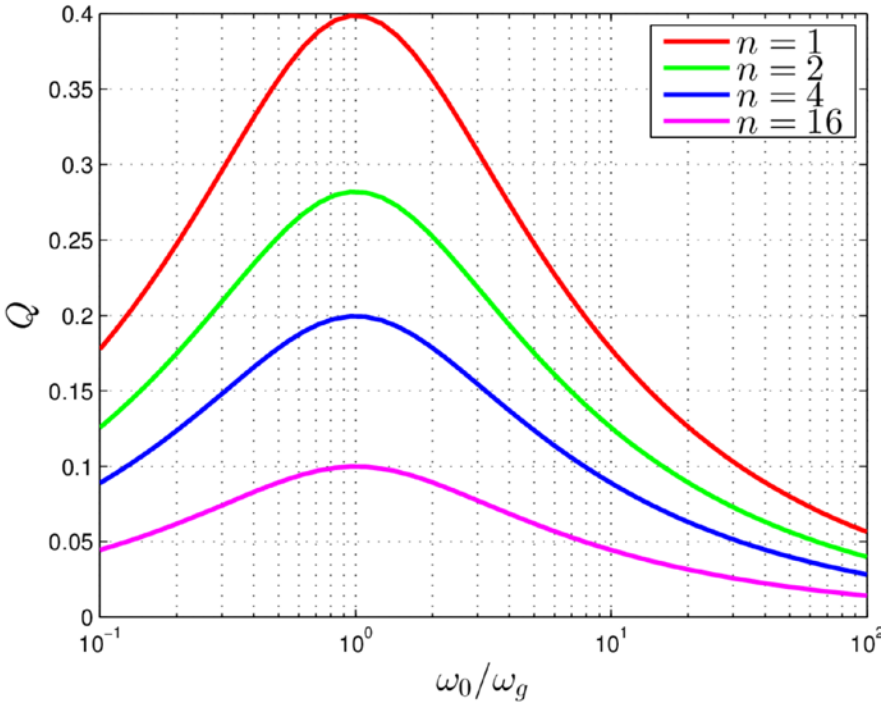
$$\Phi_{nn}(\omega) = |G_n(j\omega)|^2 \Phi_0 = \frac{\Phi_0}{1+\left(\frac{\omega}{\omega_g}\right)^2}$$

- The error of $G(j\omega)$ due to noise:

$$\sigma_{G_n} = \frac{\sqrt{2\Phi_0\omega_g}}{U_0} Q$$

- The factor Q is given by:

$$Q = \frac{\sqrt{\frac{\omega_0}{\omega_g}}}{\sqrt{\pi\left[1+\left(\frac{\omega_0}{\omega_g}\right)^2\right]}} \frac{1}{\sqrt{n}}$$



- Method is consistent in the mean square:
  - Variance of the estimate error converges to 0 as $n \to \infty$

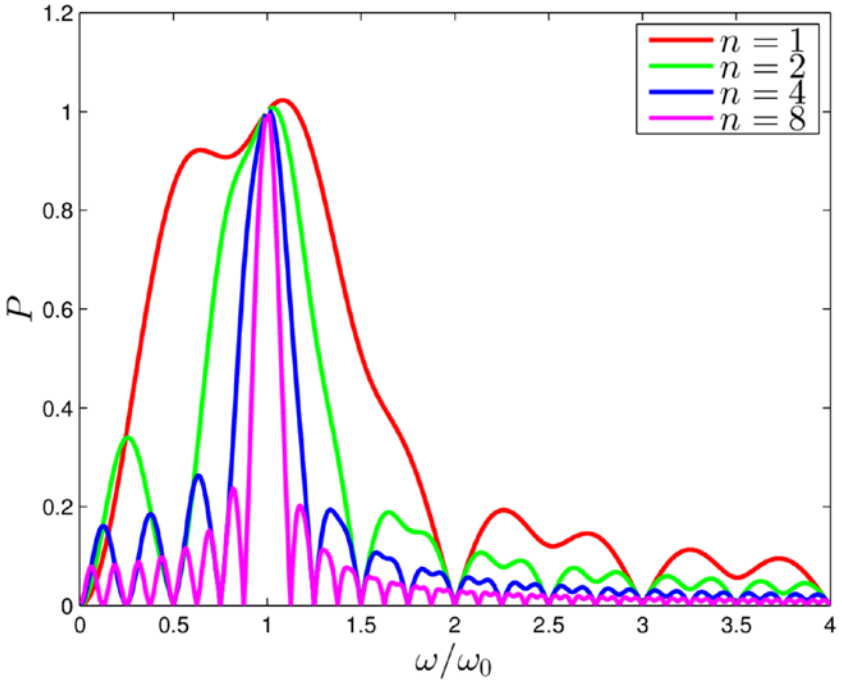
Orthogonal correlation – convergence analysis

- Sine/cosine disturbance:

$$n(t) = N_0 \cos(\omega t)$$

- The error of $G(j\omega_0)$:

$$|\Delta G(j\omega_0)| = \sigma_{G_s} = \frac{N_0}{U_0} P$$

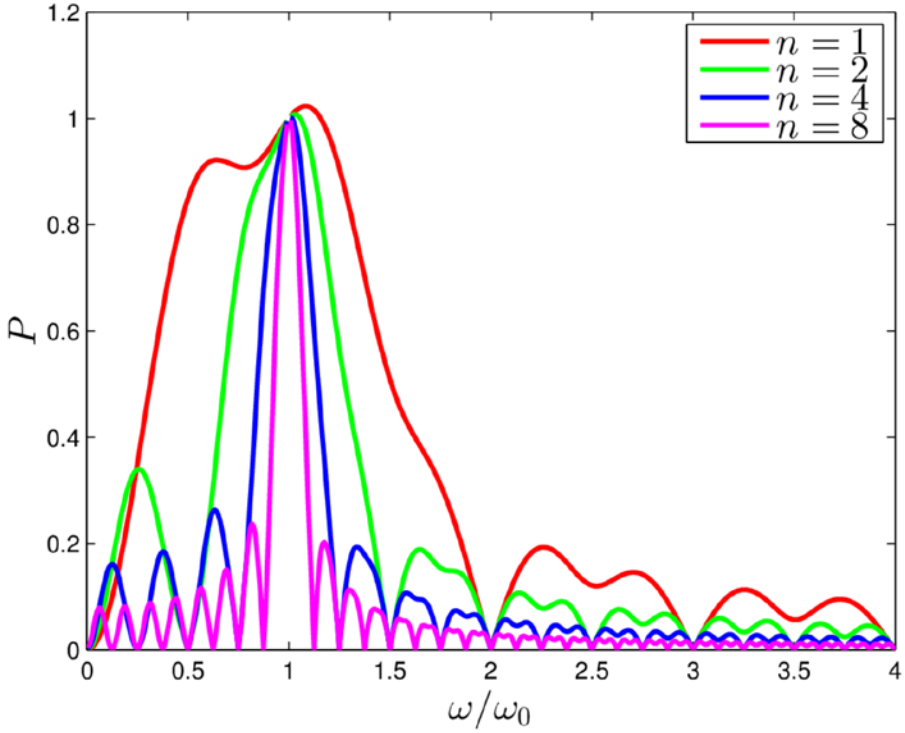


- The factor P is given by:

$$P = \frac{2\left(\frac{\omega}{\omega_0}\right)}{\left[1 - \left(\frac{\omega}{\omega_0}\right)^2\right]} \frac{\left|\sin\left(\frac{\pi n \omega}{\omega_0}\right)\right|}{\frac{\pi n \omega}{\omega_0}} \sqrt{1 - \left[1 - \left(\frac{\omega}{\omega_0}\right)^2\right] cos^2\left(\frac{\pi n \omega}{\omega_0}\right)}$$

  - The factor P has zeros at $\omega/\omega_0 = 0,2,3,4, \ldots$
    - If disturbance frequency is a multiple of a measured frequency (but not the same) or disturbance is DC, the estimate error is 0

Orthogonal correlation – convergence analysis



- The last observation holds even for finite measurement times
- Periodic disturbances of other frequencies cause an error of the frequency-response estimate in a finite measurement time nT_p:
  - The error is proportional to P
- The largest error is caused by harmonic signals with frequencies in the vicinity of the measured frequency ω_0
- Standard deviation of the error of the frequency-response estimate converges to 0 for long measurement times ($n \to \infty$):
  - Method is consistent in the mean square

Orthogonal correlation – convergence analysis

- Low-frequency non-stationary disturbances (drift):

$$n(t) = d(t)$$

$$\sigma_{G_d} = \frac{4D_0}{U_0\ \omega_0}$$

- The error of the frequency-response estimate due to drift does not depend on measurement time:
  - The error persists for long measurement times – the approach is not consistent in the mean square
- The estimate is bias-free only if the mean value of the disturbance (its mathematical expectation) is 0
- Consequently, special care is needed if drift is present in the signals (e.g., filtering)

Orthogonal correlation – non-harmonic input

- Instead of harmonic excitation, rectangular or trapezoidal input pulses can be used:
  - The implementation is usually simpler for the latter two:
    - Rectangular – a relay implements ON-OFF excitation
    - Trapezoidal – three-state controller in combination with a motor
- The idea – the input is written as a Fourier series
- The test signal is a series of rectangular pulses:
  - Fourier series expansion of the input:

$$u(t) = \frac{4}{\pi} U_0 \left[\sin(\omega_0 t) + \frac{1}{3}\sin(3\omega_0 t) + \frac{1}{5}\sin(5\omega_0 t) + \cdots \right]$$

  - The corresponding output of the linear system (freq. resp. $G(j\omega)$):

$$y(t) = \frac{4}{\pi} U_0 \left[|G(j\omega_0)| \sin[\omega_0 t + \varphi(\omega_0)] + \frac{1}{3}|G(3j\omega_0)| \sin[3\omega_0 t + \varphi(3\omega_0)] + \frac{1}{5}|G(5j\omega_0)| \sin[5\omega_0 t + \varphi(5\omega_0)] + \cdots \right]$$

Orthogonal correlation – rectangular pulses

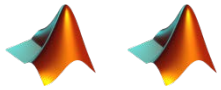
- The output signal is decomposed into a harmonic signal of the measured frequency (a useful signal in this case) and its odd multiples (disturbances in this case):
  - Note that harmonic output disturbances of a multiple of the basic frequency ω_0 do not cause any error in frequency response estimate (see factor P – link)
  - Consequently, the excitation with rectangular pulses of amplitude U_0 is equivalent to harmonic excitation of amplitude $\frac{4}{\pi}U_0$
  - The formulas are the same as in the case of harmonic excitation but multiplied with $\pi/4$:

$$\Re[G(j\omega_0)] = \frac{\pi}{2U_0 nT_p}\int_0^{nTp} y(t)\cos(\omega_0 t)dt$$

$$\Im[G(j\omega_0)] = \frac{-\pi}{2U_0 nT_p}\int_0^{nT_p} y(t)\sin(\omega_0 t)dt$$

Orthogonal correlation – rectangular pulses

- Convergence analysis (bias, consistency in the mean square) is the same as in the case of harmonic excitation:
  - Only input amplitude $\frac{4}{\pi}U_0$ has to be assumed instead of U_0
- The advantages of rectangular pulses excitation:
  - Implementation as already mentioned
  - Actuator can have extremely nonlinear static curve or characteristic
  - If compared to any other periodic input signal, the rectangular one achieves the highest amplitude of the basic harmonic frequency at fixed amplitude of the signal $\rightarrow$ the error of the estimated frequency response is therefore the lowest
  - 3rd and 5th harmonic frequency can also be estimated but not without errors



Orthogonal correlation – trapezoidal pulses

- In case of integral actuators (e.g. motors), rectangular inputs are not possible to implement, but trapezoidal ones are:

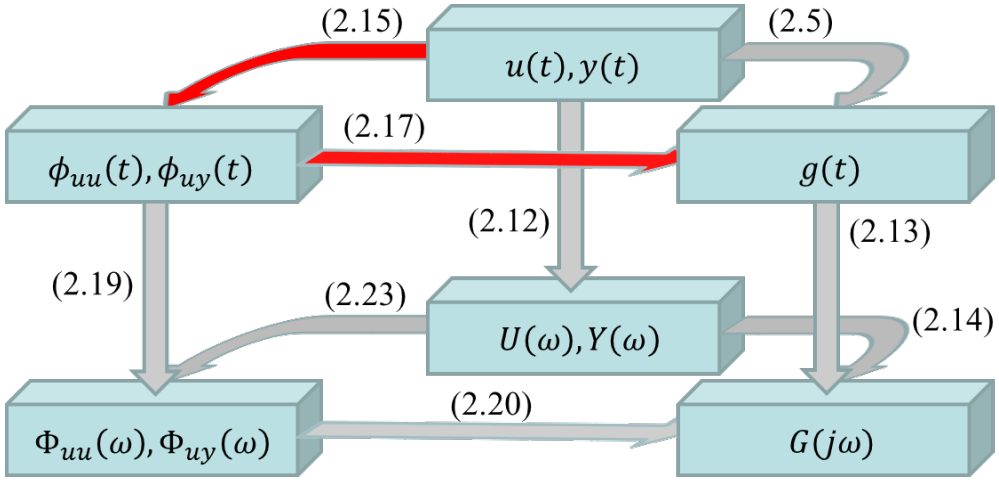
$$u(t) = \begin{cases} \dfrac{U_0 t}{T_1} & \text{for} \quad 0 \le t \le T_1 \\ U_0 & \text{for} \quad T_1 \le t \le T - T_1 \\ \dfrac{U_0 (T - t)}{T_1} & \text{for} \quad T - T_1 \le t \le T \end{cases}$$

- Fourier series expansion of the input:

$$u(t) = \frac{2U_0 T}{\pi^2 T_1}\left[\sin(\omega_0 T_1)\sin(\omega_0 t) + \frac{1}{3^2}\sin(3\omega_0 T_1)\sin(3\omega_0 t) + \right.$$

Correlation analysis

- The most direct path from time signals to non-parametric models is deconvolution ((2.5) in the figure), but it is not used in practice because of its sensitivity to noise
- A more applicable approach is correlation analysis:
  - It is also based on deconvolution
  - But in this case deconvolution is performed between cross-correlation between input and output, and auto-correlation of the input ((2.15), (2.17))
  - The result is impulse response of the process $g(\tau)$
- Frequency response $G(j\omega)$ can be obtained by applying Fourier transform on the impulse response



Correlation analysis (continuous version)

- The deconvolution operation is determination of an unknown impulse response $g(\tau)$ from the convolution equation where correlation functions $\phi_{uu}(\tau)$ and $\phi_{uy}(\tau)$ are assumed known:

$$\phi_{uy}(\tau) = \int_0^\infty g(t)\phi_{uu}(\tau - t)dt$$

  - In case of continuous-time signals this is a very complex mathematical operation that can be solved only approximatively
- Alternative approach relies on discrete signals and will be dealt with later
- The third possibility is to excite the process with white noise (high-frequent noise with very high cut-off frequency):
  - Auto-correlation function (ACF) of the white noise with power spectral density Φ_{u0} is delta-impulse:

$$\phi_{uu}(\tau) = \Phi_{u0}\delta(\tau)$$

Correlation analysis (continuous version)

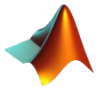
- Convolution equation is simplified if $\phi_{uu}(\tau)$ is taken into account:

$$\phi_{uy}(\tau) = \int_0^\infty g(t)\Phi_{u0}\delta(\tau - t)dt \; = g(\tau)\Phi_{u0}$$

  - Obtaining the final result is then a trivial operation:

$$g(\tau) = \frac{1}{\Phi_{u0}}\phi_{uy}(\tau)$$

- Identification procedure is very simple:
  - The process is excited with (practically) white noise – high-frequent
  - Cross-correlation function (CCF) $\phi_{uy}(\tau)$ is estimated
  - CCF $\phi_{uy}(\tau)$ is divided by input noise PSD Φ_{u0}
  - This result can be converted to frequency response by applying Fourier transform on it



Correlation analysis (convergence of CCF)

- CCF of two continuous-time stochastic signals $u(t)$ and $y(t)$ is defined as:

$$\phi_{uy}(\tau) = E\{u(t)y(t+\tau)\} =$$
$$= \lim_{T_i \to \infty} \frac{1}{T_i} \int_0^{T_i} u(t)y(t+\tau)dt = \lim_{T_i \to \infty} \frac{1}{T_i} \int_0^{T_i} u(t-\tau)y(t)dt$$

- In practice infinite time intervals cannot be used because of finite observations times T_i
- First the effect of finite observation time T_i to errors in CCF will be studied

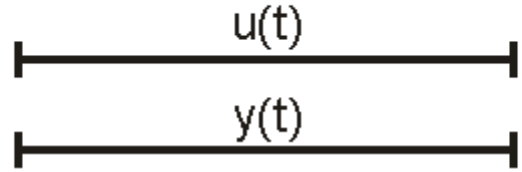
Convergence of CCF – bias (no noise)

- We assume:
  - Signals $u(t)$ and $y(t)$ are given in time interval:

$$0 \le t \le T + \tau$$

  - Their mean value is 0:

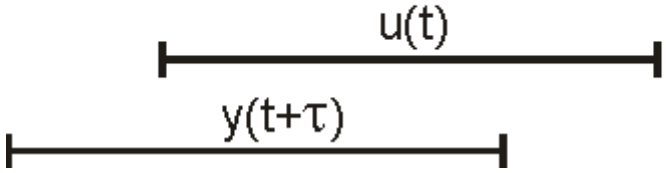
$$E\{u(t)\} = 0 \quad \text{in} \quad E\{y(t)\} = 0$$



- CCF estimate:

$$\hat{\phi}_{uy}(\tau) = \frac{1}{T}\int_0^T u(t)y(t+\tau)dt$$

$$= \frac{1}{T}\int_0^T u(t-\tau)y(t)dt$$



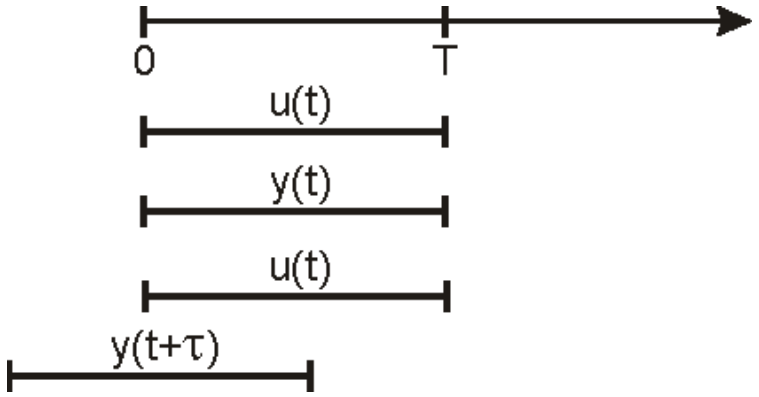
- Mathematical expectation of this estimate:

$$E\{\hat{\phi}_{uy}(\tau)\} = \frac{1}{T}\int_0^T E\{u(t)y(t+\tau)\}dt = \frac{1}{T}\int_0^T \phi_{uy}(\tau)dt = \phi_{uy}(\tau)$$

- The estimate is bias-free even in the case of finite T

Convergence of CCF – bias (no noise)

- What if mean values of $u(t)$ and $y(t)$ are not 0 and they are defined only on the time interval $0 \leq t \leq T$?
- The question is: How to treat signals in the interval where they are not defined (the argument of $y(t+\tau)$ can exceed T)?
  - The obvious choice is to substitute both signals with 0 in this case:



    - Then $\hat{\phi}_{uy}(\tau)$ for $\tau \geq 0$ becomes:

$$\hat{\phi}_{uy}(\tau) = \frac{1}{T}\int_{0}^{T-\tau} u(t)y(t+\tau)dt$$

    - Then $\hat{\phi}_{uy}(\tau)$ for $\tau \leq 0$ becomes:

$$\hat{\phi}_{uy}(\tau) = \frac{1}{T}\int_{-\tau}^{T} u(t)y(t+\tau)dt$$

Convergence of CCF – bias (no noise)

- Mathematical expectation of the estimate:
  - for $\tau \geq 0$:

$$E\{\hat{\phi}_{uy}(\tau)\} = \frac{1}{T}\int_0^{T-\tau} E\{u(t)y(t+\tau)\}dt = \frac{T-\tau}{T}\phi_{uy}(\tau)$$

  - for $\tau \leq 0$:

$$E\{\hat{\phi}_{uy}(\tau)\} = \frac{1}{T}\int_{-\tau}^{T} E\{u(t)y(t+\tau)\}dt = \frac{T+\tau}{T}\phi_{uy}(\tau)$$

- The above two equations can be rewritten as one:

$$E\{\hat{\phi}_{uy}(\tau)\} = \left[1 - \frac{|\tau|}{T}\right]\phi_{uy}(\tau)$$

  - The estimate is biased
  - The estimate is consistent, because the bias decreases with increasing observation time T; the bias diminishes as $T \to \infty$

Convergence of CCF – bias (no noise)

- If unbiased estimate is required, the formulas for $\hat{\phi}_{uy}(\tau)$ should be slightly modified – the denominator should include actual integration time instead of T:

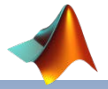
$$\hat{\phi}_{uy}(\tau) = \frac{1}{T-\tau}\int_0^{T-\tau} u(t)y(t+\tau)dt = \frac{1}{T-|\tau|}\int_0^{T-|\tau|} u(t)y(t+\tau)dt \quad \text{for } \tau \geq 0$$

$$\hat{\phi}_{uy}(\tau) = \frac{1}{T+\tau}\int_{-\tau}^{T} u(t)y(t+\tau)dt = \frac{1}{T-|\tau|}\int_{|\tau|}^{T} u(t)y(t+\tau)dt \quad \text{for } \tau \leq 0$$

- Mathematical expectation of this estimate:

$$E\{\hat{\phi}_{uy}(\tau)\} = \phi_{uy}(\tau)$$

  - The estimate is bias-free
  - But in the above expressions, the time integration is very short; moreover, the integral result is divided by $(T-|\tau|)$ which leads to very high variance of the CCF-estimate error if τ is large
  - Consequently: this estimation has to be avoided



Convergence of CCF – consistency in the mean square (no noise)

- Variance of the CCF-estimate error:

$$\mathrm{var}\left[\Delta\hat{\phi}_{uy}(\tau)\right] = \mathrm{var}\left[\hat{\phi}_{uy} - \phi_{uy}\right] = E\{[\hat{\phi}_{uy} - \phi_{uy}]^2\}$$
$$= E\{\hat{\phi}_{uy}^2 - 2\hat{\phi}_{uy}\phi_{uy} + \phi_{uy}^2\} = E\{\hat{\phi}_{uy}^2\} - \phi_{uy}^2$$

- After a lengthy derivation the following result is obtained:

$$\mathrm{var}\left[\Delta\hat{\phi}_{uy}(\tau)\right] = \frac{1}{T}\int_{-T}^{T}\left[1 - \frac{|\xi|}{T}\right]\left[\phi_{uu}(\xi)\phi_{yy}(\xi) + \phi_{uy}(\xi+\tau)\phi_{yu}(\xi-\tau)\right]d\xi$$

- If the integral exists (is finite), then it is easy to show:

$$\lim_{T\to\infty}\mathrm{var}\left[\Delta\hat{\phi}_{uy}(\tau)\right] = 0$$

  - The integral is finite if either of conditions is fulfilled:
    - $E\{u(t)\} = 0$ or
    - $E\{y(t)\} = 0$
- The estimate of CCF is then consistent in the mean square

Convergence of CCF – consistency in the mean square (no noise)

$$\mathrm{var}\left[\Delta\hat{\phi}_{uy}(\tau)\right] = \frac{1}{T}\int_{-T}^{T}\left[1-\frac{|\xi|}{T}\right]\left[\phi_{uu}(\xi)\phi_{yy}(\xi)+\phi_{uy}(\xi+\tau)\phi_{yu}(\xi-\tau)\right]d\xi$$

- Under the same condition the CCF-estimate error converges to 0 for large values of its argument – for large measurement times T the term $\frac{|\xi|}{T}$ can be omitted
- Note that the above variance is **not due to noise, but due to the stochastic nature of the two signals**:
  - In a finite time horizon T it is not possible to determine the stochastic correlation between two random signals without uncertainty
  - Therefore it is termed the ***intrinsic statistic uncertainty***
  - Variance of the error in $\hat{\phi}_{uy}(\tau)$ grows with a growing *conservation tendency* of the $u(t)$ and $y(t)$ that is reflected in their ACF and CCF

Convergence of CCF – effect of noise

- Besides intrinsic statistic uncertainty there is also uncertainty due to stochastic disturbances in signals $u(t)$ and $y(t)$
- Assume now that the signals $u(t)$ and $y(t)$ are corrupted with noise or some other type of disturbance:

$$y(t) = y_0(t) + n(t)$$
$$u(t) = u_0(t) + n_u(t)$$

- Assume also that these additive disturbances are zero-mean:

$$E\{n(t)\} = 0$$
$$E\{n_u(t)\} = 0$$

- And finally, assume that the disturbances are statistically independent of the useful signals:

$$E\{u_0(t)n_u(t)\} = E\{u_0(t)n(t)\} = E\{y_0(t)n_u(t)\} = E\{y_0(t)n(t)\} = 0$$

Convergence of CCF – bias (noise)

- In case of CCF, the estimate becomes:

$$\hat{\phi}_{uy}(\tau) = \frac{1}{T}\int_0^T [u_0(t) + n_u(t)][y_0(t+\tau) + n(t+\tau)]dt$$

- Mathematical expectation of this estimate:

$$E\{\hat{\phi}_{uy}(\tau)\} = \frac{1}{T}\int_0^T [E\{u_0(t)y_0(t+\tau)\} + E\{n_u(t)n(t+\tau)\}]dt =$$

$$= \phi_{u_0y_0}(\tau) + \phi_{n_un}(\tau)$$

- The estimate is therefore bias-free only if $\phi_{n_un} = 0$
  – this conditions is satisfied in the following holds:
  - Disturbances n_u and n are not correlated
  - At least one of the disturbances has zero mean

Convergence of CCF – consistency in the mean square (noise)

- Variance of the error of CCF estimate:

$$var\left[\Delta\hat{\phi}_{uy}(\tau)\right] = \frac{1}{T}\int_{-T}^{T}\left[1-\frac{|\xi|}{T}\right]\left[\phi_{u_0u_0}(\xi)\phi_{y_0y_0}(\xi)+\phi_{u_0u_0}(\xi)\phi_{nn}(\xi)+\right.$$
$$+\phi_{y_0y_0}(\xi)\phi_{n_un_u}(\xi)+\phi_{n_un_u}(\xi)\phi_{nn}(\xi)+$$
$$+\phi_{u_0y_0}(\xi+\tau)\phi_{y_0u_0}(\xi-\tau)+\phi_{u_0y_0}(\xi+\tau)\phi_{nn_u}(\xi-\tau)+$$
$$\left.+\phi_{n_un}(\xi+\tau)\phi_{y_0u_0}(\xi-\tau)+\phi_{n_un}(\xi+\tau)\phi_{nn_u}(\xi-\tau)\right]d\xi$$

- If ACFs and CCFs are integrable, the variance of the error goes to 0 as the integration time converges to infinity

- In this case and if both disturbances are not correlated, the estimate is consistent in the mean square

Convergence of CCF – conclusion

- The expressions for the calculations of correlation functions with the factor $\frac{1}{T}$ give biased, but consistent estimates
- The unbiased estimates are obtained if the factor $\frac{1}{T}$ is substituted by the factor $\frac{1}{T-|\tau|}$
- But in the latter case the variance of the error is larger and even converges to infinity as $|\tau| \to T$
- Consequently, the formulas with $\frac{1}{T}$ are used even though they produce biased estimates

Convergence of auto-correlation function

- Mathematical expressions for estimation of ACF, its bias and consistency are obtained from the expressions for CCF by substitution:

$$y(t) \to u(t)$$

- The estimate of ACF:

$$\hat{\phi}_{uu}(\tau) = \frac{1}{T}\int_0^T u(t)u(t+\tau)dt$$

- The estimate is bias-free only if signal $u(t)$ is not corrupted with noise as can be inferred from:

$$E\{\hat{\phi}_{uu}(\tau)\} = \frac{1}{T}\int_0^T [E\{u_0(t)u_0(t+\tau)\} + E\{n_u(t)n_u(t+\tau)\}] =$$

$$= \phi_{u_0u_0}(\tau) + \phi_{n_un_u}(\tau)$$

Convergence of auto-correlation function

- Variance of the error of ACF estimate :

$$\mathrm{var}\left[\Delta\hat{\phi}_{uu}(\tau)\right] = \frac{1}{T}\int_{-T}^{T}\left[1-\frac{|\xi|}{T}\right]$$
$$\left[\phi^2_{u_0u_0}(\xi) + 2\phi_{u_0u_0}(\xi)\phi_{n_un_u}(\xi) + \phi^2_{n_un_u}(\xi) + \right.$$
$$\phi_{u_0u_0}(\xi+\tau)\phi_{u_0u_0}(\xi-\tau) + \phi_{u_0u_0}(\xi+\tau)\phi_{n_un_u}(\xi-\tau) +$$
$$\left.+\phi_{n_un_u}(\xi+\tau)\phi_{u_0u_0}(\xi-\tau) + \phi_{n_un_u}(\xi+\tau)\phi_{n_un_u}(\xi-\tau)\right]d\xi$$

- The error converges to 0 as $T \to \infty$ if the ACF of the input and the ACF of the noise are integrable

- If these requirements are met, the estimate of ACF is consistent in the mean square

Variance of the impulse-response-estimate error

- Summary of the properties of correlation analysis:
  - Excite the process with white noise (high-frequent noise with very high cut-off frequency) – noise PSD is Φ_{u0}
  - The result is impulse response:

$$\hat{g}(\tau) = \frac{1}{\Phi_{u0}} \hat{\phi}_{uy}(\tau)$$

  - The formulas for $\hat{\phi}_{uy}(\tau)$ produce biased but consistent estimates
- The variance of the $\hat{g}(\tau)$-error:

$$\sigma_g^2(\tau) = \mathrm{var}[\Delta g\,(\tau)] = \frac{1}{T}\left[\int_0^\infty g^2(t)dt + \frac{\sigma_n^2}{\Phi_{u0}}\right]$$

  - The first term is due to intrinsic statistic uncertainty
  - The second term is due to white noise with variance σ_n^2 acting on $y(t)$
  - The estimate is consistent in the mean square if the impulse response is integrable

Correlation analysis with discrete signals

- Unlike with continuous-time systems, deconvolution is a relatively simple operation in discrete-time context
- The convolution equation is a summation that connects discrete CCF, discrete ACF and discrete impulse response:

$$\phi_{uy}(\tau) = \sum_{k=0}^{\infty} g(k)\phi_{uu}(\tau - k)$$

- Deconvolution – an operation where the unknown impulse response $g(k)$ is obtained from the known ACF $\phi_{uu}(\tau)$ and the known CCF $\phi_{uy}(\tau)$
- But of course, all these signals are of finite length due to finite length of measurements – the approach is therefore suitable only for processes with finite impulse response

Correlation analysis with discrete signals

- Stable processes:
  - After the transient (after a certain time $k = m$) the impulse response $g(k)$ becomes arbitrarily small
  - The impulse response can be neglected after that time, and can therefore be clipped at this time
- The deconvolution can therefore be seen as a system of equations:
  - By writing down the equation $\phi_{uy}(\tau)$ for a specific τ one linear equation is obtained
  - If this is repeated for more values of argument τ, a system of linear equations is obtained
  - In our case $\phi_{uy}(\tau)$ and $\phi_{uu}(\tau)$ are assumed known for $-N_1 \leq \tau \leq N_2$
  - The unknowns are the impulse response samples: $g(0), g(1), \ldots\ g(m)$
  - In order to solve the system at least $m + 1$ equations are needed
  - If the number of equation is larger, the system is overdetermined

Correlation analysis with discrete signals

- System of linear equations:

$$\phi_{uy}(-N_1+m)=\phi_{uu}(-N_1+m)g(0)+\phi_{uu}(-N_1+m-1)g(1)+\cdots+\phi_{uu}(-N_1)g(m)$$
$$\vdots$$
$$\phi_{uy}(-1)=\phi_{uu}(-1)g(0)+\phi_{uu}(-2)g(1)+\cdots+\phi_{uu}(-1-m)g(m)$$
$$\phi_{uy}(0)=\phi_{uu}(0)g(0)+\phi_{uu}(-1)g(1)+\cdots+\phi_{uu}(-m)g(m)$$
$$\phi_{uy}(1)=\phi_{uu}(1)g(0)+\phi_{uu}(0)g(1)+\cdots+\phi_{uu}(-m+1)g(m)$$
$$\vdots$$
$$\phi_{uy}(N_2)=\phi_{uu}(N_2)g(0)+\phi_{uu}(N_2-1)g(1)\ldots+\phi_{uu}(N_2-m)g(m)$$

- The system in the vector-matrix form: $\boldsymbol{\Phi}_{uy}=\boldsymbol{\Phi}_{uu}\mathbf{g}$

$$\boldsymbol{\Phi}_{uy}=\begin{bmatrix}\phi_{uy}(-N_1+m)\\ \vdots\\ \phi_{uy}(-1)\\ \phi_{uy}(0)\\ \phi_{uy}(1)\\ \vdots\\ \phi_{uy}(N_2)\end{bmatrix}\quad \boldsymbol{\Phi}_{uu}=\begin{bmatrix}\phi_{uu}(-N_1+m),\phi_{uu}(-N_1+m-1),\ldots,\phi_{uu}(-N_1)\\ \vdots\\ \phi_{uu}(-1),\phi_{uu}(-2),\ldots,\phi_{uu}(-1-m)\\ \phi_{uu}(0),\phi_{uu}(-1),\ldots,\phi_{uu}(-m)\\ \phi_{uu}(1),\phi_{uu}(0),\ldots,\phi_{uu}(1-m)\\ \vdots\\ \phi_{uu}(N_2),\phi_{uu}(N_2-1),\ldots,\phi_{uu}(N_2-m)\end{bmatrix}\quad \mathbf{g}=\begin{bmatrix}g(0)\\ g(1)\\ \vdots\\ g(m)\end{bmatrix}$$

Correlation analysis with discrete signals

- The number of equations: $N_1 - m + N_2 + 1$
- The number of unknowns: $m + 1$
- If

$$N_1 - m + N_2 + 1 = m + 1$$

- or equivalently

$$N_1 + N_2 = 2m$$

- the system is square and the solution is:

$$\mathbf{g} = \mathbf{\Phi}_{uu}^{-1}\mathbf{\Phi}_{uy}$$

- If ($N_1 + N_2 > 2m$), the solution is obtained by least square parameter estimation:

$$\mathbf{g} = [\mathbf{\Phi}_{uu}^{T}\mathbf{\Phi}_{uu}]^{-1}\mathbf{\Phi}_{uu}^{T}\mathbf{\Phi}_{uy}$$

Correlation analysis with discrete signals

- The continuous-time correlation analysis where the process is excited with white noise can be extended to the discrete case
- ACF of the white noise with PSD Φ_{u0}:

$$\phi_{uu}(\tau) = \Phi_{u0}\delta(\tau)$$

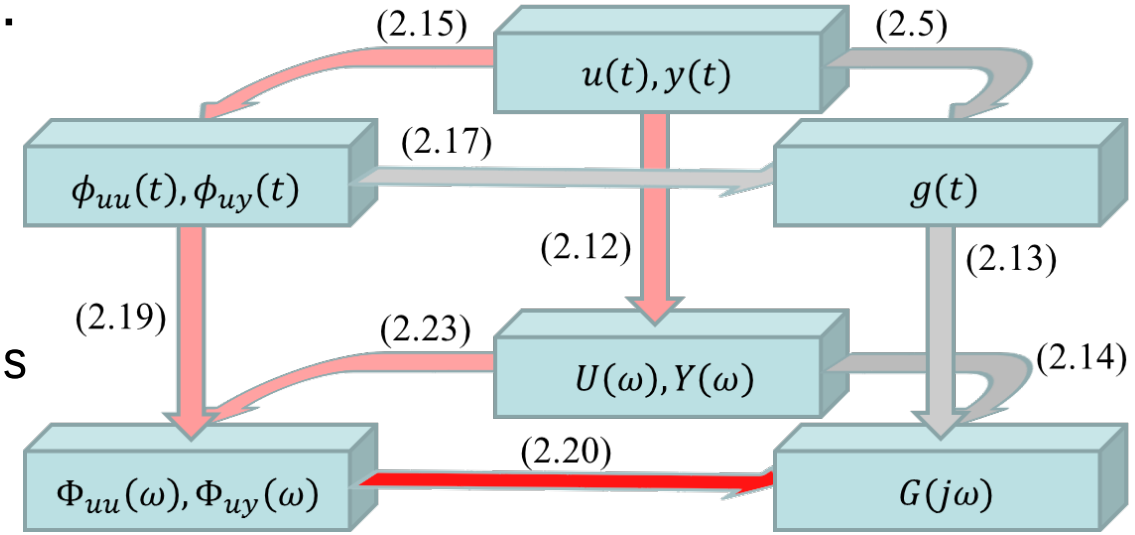
- Because of special form of $\phi_{uu}(\tau)$, the deconvolution simplifies to:

$$g(\tau) = \frac{1}{\Phi_{u0}}\phi_{uy}(\tau)$$

- Identification procedure:
  - The process is excited with input noise
  - CCF between the input and the output is calculated
  - CCF is divided with Φ_{u0} to obtain impulse response
  - The latter can be transformed with Fourier transformation to obtain frequency response if needed

Spectral analysis

- Frequency response $G(j\omega)$ can also be obtained by dividing power spectral densities $\Phi_{uy}(\omega)$ and $\Phi_{uu}(\omega)$ – path (2.20)
- PSDs can be obtained:
  - with Fourier transform of ACF $\phi_{uu}(\tau)$ and CCF $\phi_{uy}(\tau)$ – path (2.15)+(2.19)
  - from Fourier transforms $U(\omega)$ and $Y(\omega)$ – path (2.12)+(2.23)



- The example shows that both paths are equivalent, but when more periods of measurements are used, the path through correlation functions is less sensitive to noise

Spectral analysis

- But have already seen that:
  - Path (2.12) is very sensitive to noise in signals – as a result Fourier transforms $U(\omega)$ and $Y(\omega)$ also become noisy – in (2.23) this noise is directly transferred to $\Phi_{uy}(\omega)$ and $\Phi_{uu}(\omega)$, and later to $G(j\omega)$
  - Path (2.15) is problematic because correlation functions have high uncertainty for large shifts τ, but Fourier transform (2.19) puts equal gains to all samples of correlation functions
- The solutions to the above mentioned problems are:
  - Filtering of power spectral densities
  - Weighting of correlation functions according to the shift τ before calculating PSDs
- We will show that both approaches are essentially equivalent and they share the same name – spectral analysis

Spectral analysis

- First the weighting of correlation functions by window functions will be treated when calculating power spectral densities:

$$\Phi_{uy}^{W}(\omega) = \mathcal{F}[\phi_{uy}(\tau)w(\tau)] = \int_{-\infty}^{\infty} \phi_{uy}(\tau)w(\tau)e^{-j\omega\tau}d\tau$$

$$\Phi_{uu}^{W}(\omega) = \mathcal{F}[\phi_{uu}(\tau)w(\tau)] = \int_{-\infty}^{\infty} \phi_{uu}(\tau)w(\tau)e^{-j\omega\tau}d\tau$$

  - The weighting function $w(\tau)$ is called a window function and it has to fulfil the following:
    - $w(\tau) = w(-\tau)$
    - $w(\tau)|_{\tau=0} = 1$
    - $w(\tau)$ should be monotonically non-increasing for positive shifts τ
    - $w(\tau)|_{\tau\to\infty} = 0$

Spectral analysis

- The equations in the previous slide present Fourier transforms of the product of continuous-time signals
- Product in time domain is transformed into convolution in the frequency domain:

$$\Phi_{uy}^{W}(\omega_0) = \int_{-\infty}^{\infty} \Phi_{uy}(\omega) W(\omega_0 - \omega) d\omega$$

$$\Phi_{uu}^{W}(\omega_0) = \int_{-\infty}^{\infty} \Phi_{uu}(\omega) W(\omega_0 - \omega) d\omega$$

  - Fourier transform $W(\omega) = \mathcal{F}[w(t)]$ is an even function of frequency which means that $W(\omega_0 - \omega) = W(\omega - \omega_0)$, and therefore:

$$\Phi_{uy}^{W}(\omega_0) = \int_{-\infty}^{\infty} W(\omega - \omega_0) \Phi_{uy}(\omega) d\omega$$

$$\Phi_{uu}^{W}(\omega_0) = \int_{-\infty}^{\infty} W(\omega - \omega_0) \Phi_{uu}(\omega) d\omega$$

Spectral analysis

$$\Phi_{uy}^{W}(\omega_0) = \int_{-\infty}^{\infty} W(\omega - \omega_0)\Phi_{uy}(\omega)d\omega$$

$$\Phi_{uu}^{W}(\omega_0) = \int_{-\infty}^{\infty} W(\omega - \omega_0)\Phi_{uu}(\omega)d\omega$$

- The above integrals can be interpreted as calculation of power in the frequency window:
  - If $W(\omega - \omega_0)$ were 1, the complete power would be obtained
  - If, on the other hand, $W(\omega - \omega_0) = \delta(\omega - \omega_0)$, there would not be any filtering: $\Phi_{uy}^{W}(\omega_0) = \Phi_{uy}(\omega_0)$
  - However, in the case of spectral analysis, $W(\omega - \omega_0)$ is a narrow function that has large values only around $\omega = \omega_0 \rightarrow \Phi^{W}(\omega_0)$ is obtained by smoothing down the original $\Phi(\omega)$ around $\omega = \omega_0$
  - Original method of spectral analysis did what the name suggests – it used band-pass filters to obtain power in the frequency band

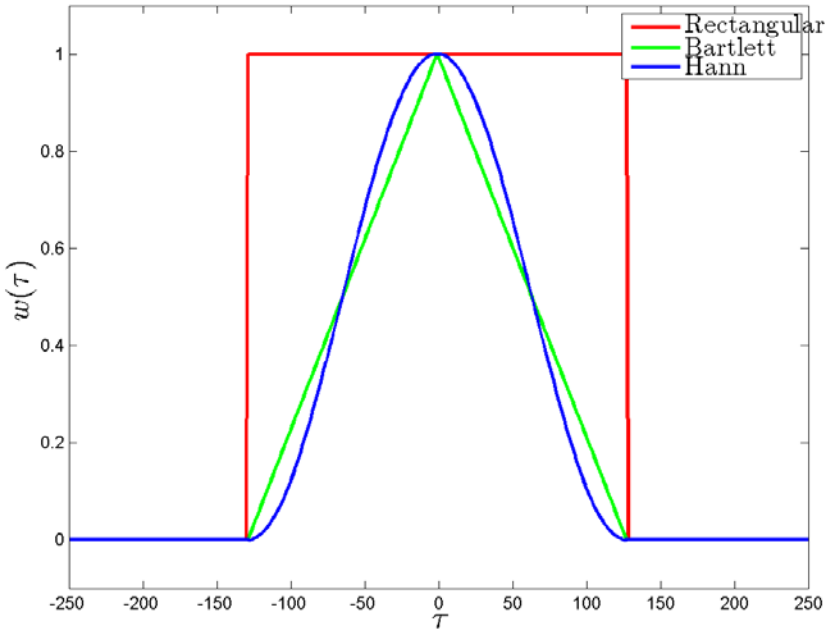
Spectral analysis

- In the literature a lot of window functions have been proposed:
  - Rectangular window:

$$w_1(\tau) = \begin{cases} 1 & |\tau| \leq T_M \\ 0 & |\tau| > T_M \end{cases}$$

  - Bartlett‘s window:

$$w_2(\tau) = \begin{cases} 1 - \frac{\tau}{T_M} & |\tau| \leq T_M \\ 0 & |\tau| > T_M \end{cases}$$



  - Hann‘s (also „Hanning‘s“) – a version of Hann‘s is Hamming‘s:

$$w_3(\tau) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\frac{\pi\tau}{T_M}\right)\right] & |\tau| \leq T_M \\ 0 & |\tau| > T_M \end{cases}$$

Spectral analysis

- Fourier transforms of these time windows:
  - Rectangular :

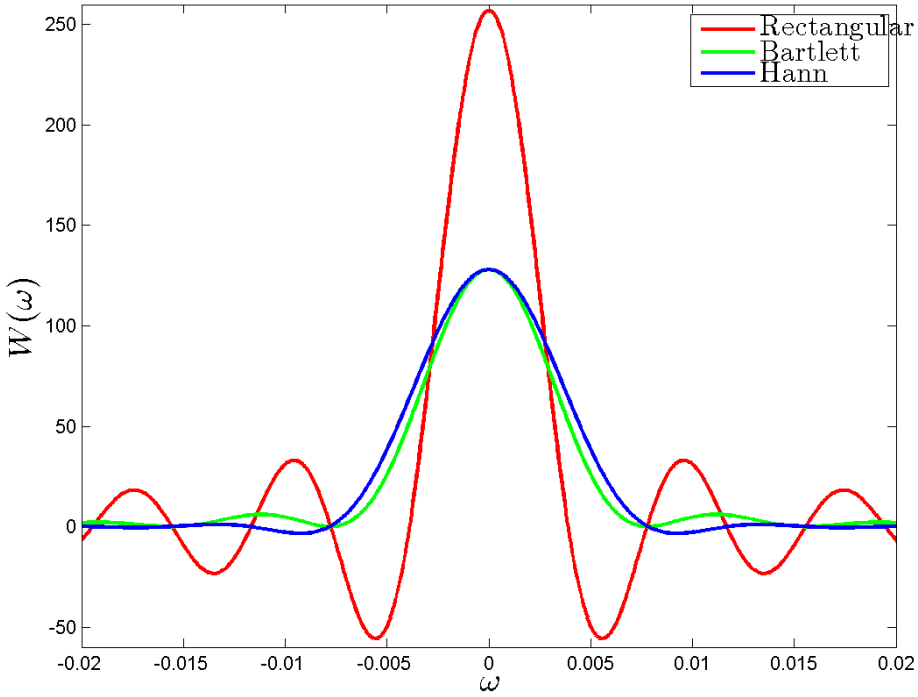
$$W_1(\omega) = 2T_M \frac{\sin(\omega T_M)}{\omega T_M}$$

  - Bartlett‘s:

$$W_2(\omega) = T_M \left(\frac{\sin\frac{\omega T_M}{2}}{\frac{\omega T_M}{2}} \right)^2$$

  - Hann‘s:

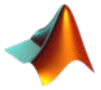
$$W_3(\omega) = T_M \frac{\pi^2 \sin(\omega T_M)}{(\pi^2 - \omega^2 T_M^2)\omega T_M}$$



Spectral analysis

Time window		Frequency window
Wide → almost „the whole“ correlation function taken into account	⇔	Narrow → PSD only slightly filtered (or averaged with neighbouring frequencies)
Narrow → the correlation function neglected at large shifts	⇔	Wide → PSD filtered considerably (mixing the neighbouring frequencies)

- Important: Filtering or averaging of PSDs introduces bias (in the part where PSDs are not constant)
- Guidelines for the choice of time window width T_M:
  - T_M small compared to the measurement time → random errors at large shifts of $\phi(\tau)$ disregarded
  - T_M large enough so that the majority of information is retained, i.e., $\phi(\tau) \ll \phi(0)$ for $\tau \geq T_M$
- Windows can filter out errors, but can also introduce new ones!



Identifiability of the process in the closed loop

- The basic questions in this context:
  Is it possible to identify the process?

- More specifically, in the area of parametric models identification:
  Is it possible to determine the process structure and estimate its parameters?

- Why posing this question?
  - The majority of identification methods requires the process input to be non-correlated with the disturbance
  - The above requirement is often violated → especially problematic is the case of closed-loop operation where process output influences process input and thus the disturbance is brought to the input

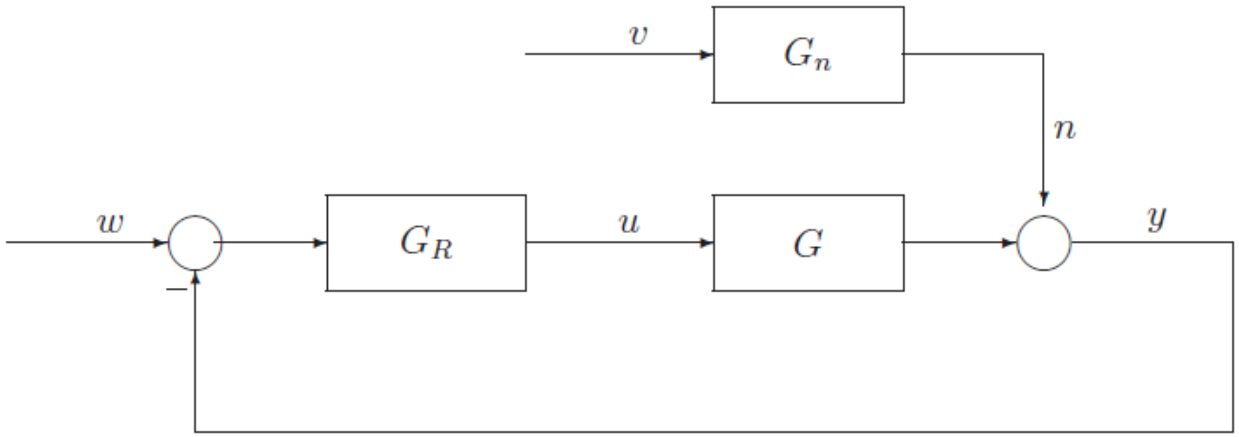
Identifiability of the process in the closed loop

- On the other hand, a closed loop often cannot be avoided:
  - In certain cases it is an inherent part of the system
  - If the process is unstable or poorly damped, the operation in the open loop is not acceptable (safety, quality)
  - Certain control systems do this normally – adaptive systems identify parameters and tune the controllers on line
- Identifiability will be analysed as follows:
  - Identifiability of non-parametric models (only for continuous models)
  - Identifiability of parametric models (only for discrete models)
  - (The path above is chosen because of clarity of presentation, and not due to some actual differences in the identifiability)
- It will be shown that the identifiability requirements are the same:
  - for both model classes (parametric and non-parametric)
  - for direct and indirect parameter estimation

Identifiability of the non-parametric models

- The structure of the closed-loop system – two external excitations (reference signal w and disturbance v):

$$w' = G_R(s)w$$
$$v' = G_R(s)G_n(s)v$$



$$y = \frac{1}{1 + G_R(s)G(s)}[G_R(s)G(s)w + G_n(s)v] = \frac{1}{1 + G_R(s)G(s)}\left[G(s)w' + \frac{1}{G_R(s)}v'\right]$$

$$u = \frac{1}{1 + G_R(s)G(s)}[G_R(s)w - G_n(s)G_R(s)v] = \frac{1}{1 + G_R(s)G(s)}[w' - v']$$

Identifiability of the non-parametric models

- v is stochastic $\rightarrow$ analysis is based on PSDs and not FT:
  - Reference w is not correlated with disturbance v
  - Consequently: w' is not correlated with v', or equiv. $\Phi_{w'v'}(\omega) = 0$
  - Therefore:

$$\Phi_{uu}(\omega) = \left|\frac{1}{1 + G_R(j\omega)G(j\omega)}\right|^2 [\Phi_{w'w'}(\omega) + \Phi_{v'v'}(\omega)]$$

$$\Phi_{uy}(\omega) = \left|\frac{1}{1 + G_R(j\omega)G(j\omega)}\right|^2 \left[G(j\omega)\Phi_{w'w'}(\omega) - \frac{1}{G_R(j\omega)}\Phi_{v'v'}(\omega)\right]$$

- Frequency response estimate is obtained by dividing the above expressions:

$$\hat{G}(j\omega) = \frac{\Phi_{uy}(\omega)}{\Phi_{uu}(\omega)} = \frac{G(j\omega)\Phi_{w'w'}(\omega) - \frac{1}{G_R(j\omega)}\Phi_{v'v'}(\omega)}{\Phi_{w'w'}(\omega) + \Phi_{v'v'}(\omega)}$$

Identifiability of the non-parametric models

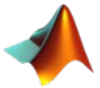
- Rewriting the last equation:

$$\hat{G}(j\omega) = \frac{\Phi_{w'w'}(\omega)}{\Phi_{w'w'}(\omega)+\Phi_{v'v'}(\omega)} G(j\omega) + \frac{\Phi_{v'v'}(\omega)}{\Phi_{w'w'}(\omega)+\Phi_{v'v'}(\omega)} \left(-\frac{1}{G_R(j\omega)}\right)$$

  - The identified frequency response is a weighted sum of the actual one and the one obtained from the inverse model of the controller

- $\hat{G}(j\omega)$ is bias-free at a particular frequency ω if:
  - This frequency is excited by the reference ($\Phi_{w'w'}(\omega) > 0$) and
  - This frequency is not excited by the disturbance ($\Phi_{v'v'}(\omega) = 0$)

- In the absence of the reference signal excitation ($\Phi_{w'w'}(\omega) = 0$), the identified model is the inverse of the controller ($-1/G_R(j\omega)$)



Identifiability of the parametric models

- Parametric models (structure and parameters) also cannot be determined uniquely in the absence of the reference excitation
- Let‘s prove this using the following example :
  - The proces with input u and output y is disturbed with a noise obtained by filtering white noise v through $G_n(z) = \frac{D(z^{-1})}{A(z^{-1})}$:

$$y = G(z)u + G_n(z)v = \frac{B(z^{-1})}{A(z^{-1})}z^{-d}u + \frac{D(z^{-1})}{A(z^{-1})}v$$

  - A feedback controller $\frac{Q(z^{-1})}{P(z^{-1})}$ is introduced as:

$$u = G_R(w - y) = \frac{Q(z^{-1})}{P(z^{-1})}(w - y)$$

  - It is assumed that there is no reference ($w = 0$):

$$y = \frac{B(z^{-1})}{A(z^{-1})}z^{-d}\left(-\frac{Q(z^{-1})}{P(z^{-1})}y\right) + \frac{D(z^{-1})}{A(z^{-1})}v$$

Identifiability of the parametric models

- The equation is multiplied with $A(z^{-1})$, the model $v \to y$ becomes:

$$\left[A(z^{-1}) + B(z^{-1})z^{-d}\frac{Q(z^{-1})}{P(z^{-1})}\right]y = D(z^{-1})v$$

- An arbitrary polynomial $S(z^{-1})z^{-d}$ can be added and subtracted:

$$\left[A(z^{-1}) + S(z^{-1})z^{-d} + B(z^{-1})z^{-d}\frac{Q(z^{-1})}{P(z^{-1})} - S(z^{-1})z^{-d}\right]y = D(z^{-1})v$$

$$\left[A(z^{-1}) + S(z^{-1})z^{-d} + \left(B(z^{-1}) - \frac{P(z^{-1})}{Q(z^{-1})}S(z^{-1})\right)z^{-d}\frac{Q(z^{-1})}{P(z^{-1})}\right]y = D(z^{-1})v$$

$$\left[Q(z^{-1})[A(z^{-1}) + S(z^{-1})z^{-d}] + [Q(z^{-1})B(z^{-1}) - P(z^{-1})S(z^{-1})]z^{-d}\frac{Q(z^{-1})}{P(z^{-1})}\right]y = Q(z^{-1})D(z^{-1})v$$

- The coloured transfer functions in the last equation are denoted by $A^*(z^{-1})$, $B^*(z^{-1})$, and $D^*(z^{-1})$:
  - In the absence of excitation ($w = 0$), the response of the original closed-loop system (process $\frac{B(z^{-1})}{A(z^{-1})}z^{-d}$ and noise filter $\frac{D(z^{-1})}{A(z^{-1})}$) to noise v is the same as obtained by process $\frac{B^*(z^{-1})}{A^*(z^{-1})}z^{-d}$ and noise filter $\frac{D^*(z^{-1})}{A^*(z^{-1})}$

Identifiability of the parametric models

- Conclusions of the example:
  - Extended process (given by polynomials $A^*(z^{-1})$, $B^*(z^{-1})$, and $D^*(z^{-1})$) is of higher order than the original process, but the response of both systems is the same
  - **Higher order is not due to pole-zero combinations that cancel each other!**
    - An identical behaviour (response to v) can be obtained by a higher-order system where the poles do not cancel with zeros
  - The structure (the order) **cannot** be determined uniquely
  - The lowest order of the extended transfer function is obtained if the order of $S(z^{-1})$ is 0 $\rightarrow$ the extended transfer function is the same as the original one
  - If the reference excitation is absent, the knowledge of the transfer function order is necessary for identifiability

Identifiability of the parametric models – parameter estimation

- Next, the requirements for parameter estimation in the absence of reference excitation will be established:
  - The only excitation is the noise v (white by assumption)
  - The new (closed-loop) noise filter is:

$$G_n'(z) = \frac{Y(z)}{V(z)} = \frac{G_n(z)}{1+G_{\mathrm{R}}(z)G(z)}$$

  - Process $G(z)$, noise filter $G_n(z) \rightarrow$ unknown with known structure:

$$G(z) = \frac{b_1 z^{-1}+\cdots+b_{n_b} z^{-n_b}}{1+a_1 z^{-1}+\cdots+a_{n_a} z^{-n_a}} z^{-d} \qquad G_n(z) = \frac{1+d_1 z^{-1}+\cdots+d_{n_d} z^{-n_d}}{1+a_1 z^{-1}+\cdots+a_{n_a} z^{-n_a}}$$

  - The controller $G_R(z) \rightarrow$ known:

$$G_R(z) = \frac{Q(z^{-1})}{P(z^{-1})} = \frac{q_0+q_1 z^{-1}+\cdots+q_\nu z^{-\nu}}{1+p_1 z^{-1}+\cdots+p_\mu z^{-\mu}}$$

  - The new noise filter is rewritten as:

$$G_n' = \frac{G_n(z)}{1+G_R(z)G(z)} = \frac{D(z^{-1})P(z^{-1})}{A(z^{-1})P(z^{-1})+B(z^{-1})z^{-d}Q(z^{-1})} = \frac{1+\delta_1 z^{-1}+\cdots+\delta_r z^{-r}}{1+\alpha_1 z^{-1}+\cdots+\alpha_\ell z^{-\ell}} = \frac{\mathcal{D}(z^{-1})}{\mathcal{A}(z^{-1})}$$

Identifiability of the parametric models – parameter estimation

- The first interpretation of the identifiability:
  Parametric model is identifiable if the parameters of $G(z)$ and $G_n(z)$ can be obtained (uniquely) from the identified $G'_n(z)$
- Identification of $G'_n(z)$ – extended LS (ELS) parameter estimation:
  - Normally $A(z^{-1})$, $B(z^{-1})$, $D(z^{-1})$ are estimated by ELS
  - Here, the excitation only comes from noise:
    - only parameters of $\mathcal{A}(z^{-1})$ and $\mathcal{D}(z^{-1})$ are estimated
    - only y and e are in the regressor vector $\boldsymbol{\psi}$
  - ELS gives the estimates: coefficients $\alpha_1 \dots \alpha_\ell$ and $\delta_1 \dots \delta_r$
  - Known values: controller coefficients $q_0 \dots g_\nu$ and $p_1 \dots p_\mu$
  - The system of equations is obtained from the equation for $G'_n(z)$:

$$D(z^{-1})P(z^{-1}) = \mathcal{D}(z^{-1})$$
$$A(z^{-1})P(z^{-1}) + B(z^{-1})z^{-d}Q(z^{-1}) = \mathcal{A}(z^{-1})$$

  - Solving the system we get: $b_1, \dots, b_{n_b}, a_1 \ \dots, a_{n_a}$, and $d_1, \dots, d_{n_d}$

Identifiability of the parametric models – parameter estimation

- But we have to check the conditions for existence of unique solutions
- Unknown coefficients can be determined if the number of equations higher or equal to the number of unknown parameters:
  - Number of equations = order of polynomials that are set equal
  - Number of unknowns = number of unknown coefficients
  - Determination of $D(z^{-1})$ from $\mathcal{D}(z^{-1})$: no problems because always more equations ($r = n_d + \mu$) than unknowns (n_d)
  - Determination of a_i and b_i from $\mathcal{A}\ (z^{-1})$:

    $$A(z^{-1})P(z^{-1}) + B(z^{-1})z^{-d}Q(z^{-1}) = \mathcal{A}(z^{-1})$$

    - Number of equations: $\text{order}\{AP + Bz^{-d}Q\} = \max\{n_a + \mu, n_b + \text{d} + \nu\}$
    - Number of unknowns: $n_a + n_b$
    - a_i and b_i can be determined if $max\{n_a + \mu, n_b + d + \nu\} \geq n_a + n_b$
- Parameter identifiability requirement for the known order of the unknown transfer function in the closed-loop where the reference excitation is absent (the above condition is rewritten):

$$max\{\mu - n_b, \nu + d - n_a\} \geq 0$$

Identifiability of the parametric models – parameter estimation

- The method just presented is an indirect one:
  - First the closed-loop noise filter $G_n'(z)$ was estimated
  - Process $G(z)$ was derived from the identified $G_n'(z)$
- Now a direct method will be treated – process coefficients will be estimated directly from measurements of u and y:
  - Note that again the only excitation comes from white noise v
  - In each sampling instant the regressor vector $\boldsymbol{\psi}^T(k)$ is constructed:

$$[-y(k-1), \dots, -y(k-n_a), u(k-d-1), \dots, u(k-d-n_b), e(k-1), \dots, e(k-n_d)]$$

  - Matrix $\boldsymbol{\Psi}$ is constructed from these row vectors
  - The parameters can be obtained, i.e., the solution exists, if the columns of $\boldsymbol{\Psi}$ are linearly independent:
    - But the control signal is obtained applying transfer function $\frac{Q(z^{-1})}{P(z^{-1})}$ on the control error (in the absence of the reference it reduces to $(-y)$):
      $u(k) = -q_0 y(k) - \dots - q_\nu y(k-\nu) - p_1 u(k-1) - \dots - p_\mu u(k-\mu)$

Identifiability of the parametric models – parameter estimation

- The following question should be posed: How to choose the controller parameters to achieve linear independence in $\boldsymbol{\Psi}$?
  - Linear dependence happens if some signal from $\boldsymbol{\psi}^T(k)$ is a linear combination of signals that are already in $\boldsymbol{\psi}^T(k)$
  - Linear dependence is avoided if each input in $\boldsymbol{\psi}^T(k)$ depends on at least one signal that is older than all the signals already in $\boldsymbol{\psi}^T(k)$
  - The most critical is the most recent input, i.e. $u(k-d-1)$:
    $u(k-d-1) = -q_0 y(k-d-1) - \cdots - q_\nu y(k-d-\nu-1) -$
    $- p_1 u(k-d-1) \ldots - p_\mu u(k-d-\mu-1)$
  - Lin. independence $\rightarrow$ either of the red signals is older than those from $\boldsymbol{\psi}$:

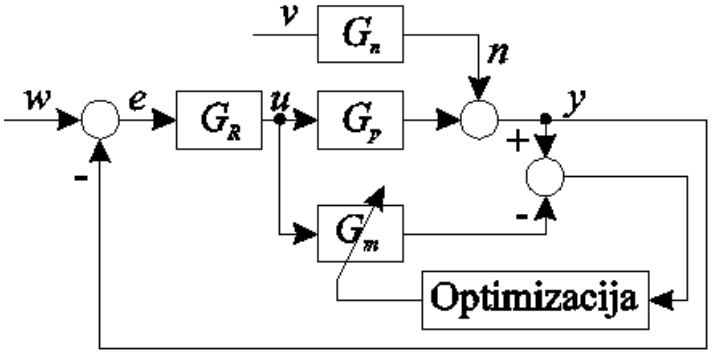
    OR

    - The oldest output in $\boldsymbol{\psi}^T(k)$ is $y(k-n_a)$, but $y(k-d-\nu-1)$ is older:
      $k-d-\nu-1 \leq k-n_a-1 \Leftrightarrow \nu+d-n_a \geq 0$
    - The oldest input in $\boldsymbol{\psi}^T(k)$ is $u(k-d-n_b)$, but $u(k-d-\mu-1)$ is older:
      $k-d-\mu-1 \leq k-d-n_b-1 \Leftrightarrow \mu-n_b \geq 0$
  - Both conditions can be combined with logical disjunction:

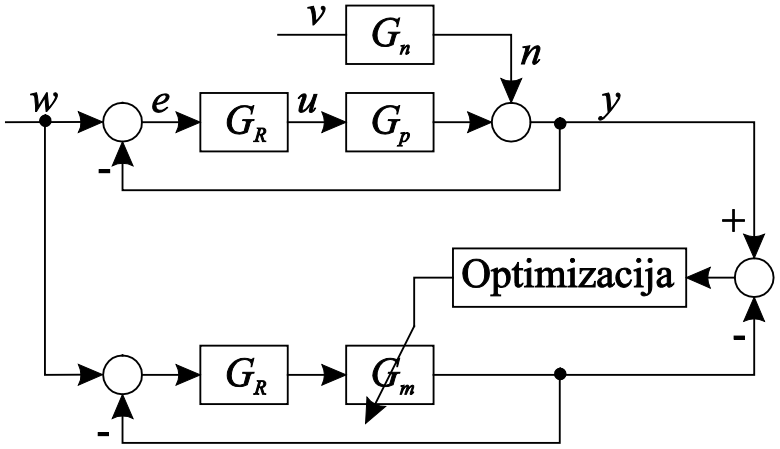
$\max\{\mu-n_b, \nu+d-n_a\} \geq 0$ (the same conditions as with indirect estimation)

Identifiability of the parametric models – parameter estimation

- How to avoid linear dependence of signals in $\boldsymbol{\psi}$ and therefore of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ in case of the excitation by the reference ?
  - By including at least as many frequencies as there are unknowns
  - Non-singularity of matrix $\boldsymbol{\Psi}^T\boldsymbol{\Psi}$ enables the existence of solution...
  - ... **but this does not guarantee bias-free estimates**
- How to achieve bias-free estimates?
  - In open-loop setup – input and disturbance not correlated
  - In closed-loop setup?



Open-loop vs.



Closed-loop model tuning

Matlab examples

- Matlab\Uvod\Primer01.m
- Matlab\Uvod\Primer02.m
- Matlab\Uvod\Primer03.m
- Matlab\Uvod\Primer04.m
- Matlab\Uvod\Primer05.m
- Matlab\Uvod\Primer06.m
- Matlab\Uvod\y_je_konv_u_g.m
- Matlab\Strejc\Primer01_P_proces_.m
- Matlab\Strejc\Primer02_I_proces_.m
- Matlab\Prilagajanje\primer1.m
- Matlab\Prilagajanje\primer2.m
- Matlab\Prilagajanje\primer3.m
- Matlab\Prilagajanje\primer4_1.m
- Matlab\Prilagajanje\primer4_2.m
- Matlab\Sumi\Primer01_Distribucija.m
- Matlab\Sumi\Primer02_Avtokor_Belega.m
- Matlab\Sumi\Primer03_Avtokor_Barvnega_Gauss.m
- Matlab\Sumi\Primer04_Avtokor_Barvnega_enakom.m
- Matlab\Sumi\Primer05_Avtokor_Barvnega_binar.m
- Matlab\Sumi\Primer06_Poissonov_sum.m
- Matlab\Sumi\Primer07_Zvezni_NBS.m
- Matlab\Sumi\Primer08_Diskretni_NBS.m
- Matlab\Sumi\Primer09_Diskret_PNBS.m
- Matlab\Pristranskost\Primer01_Konsistenca_ocene_sr_vred.m
- Matlab\Pristranskost\Primer02_Pristranskost_ocene_var.m
- Matlab\MNK\Primer01_Skalarni_problem.m
- Matlab\MNK\Primer02_Vektorski_problem.m
- Matlab\MNK\Primer03_Bazne_funkcije.m
- Matlab\MNK\Primer04_Dinamicni_sis_nerekurzivno.m
- Matlab\MNK\Primer05_Dinamicni_sis_rekurzivno.m
- Matlab\MNK\Primer06_Lastne_vred_vekt_Singular_vred.m
- Matlab\MNK\Primer07_Faktor_pogojenosti.m
- Matlab\MNK_Mod\Primer01_RNK_beli_sum.m
- Matlab\MNK_Mod\Primer02_RNK_poz_real.m
- Matlab\MNK_Mod\Primer03_RNK_1_A.m
- Matlab\MNK_Mod\Primer04_NK_zvezno_majhen_sum.m
- Matlab\MNK_Mod\Primer05_NK_zvezno_sred_sum.m
- Matlab\MNK_Mod\Primer06_NK_zvezno_velik_sum.m
- m-datoteke-Pogl_3_1\Four_trans_signalov.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_1.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_2.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_3.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_4.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_5.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_6.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_7.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_8.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_9.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_10.m
- m-datoteke-Pogl_3_1\ID_Fourier_analiza_primer_11.m
- m-datoteke-Pogl_3_2\ID_Analiza_frekvencnega_odziva_Primer1.m
- m-datoteke-Pogl_3_2\ID_Analiza_frekvencnega_odziva_Primer2.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Primer_1.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Primer_2.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Primer_3.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Primer_4.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Primer_5.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Primer_6.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Primer_7.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Primer_8.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Pravokotni_primer1.m
- m-datoteke-Pogl_3_2\ID_Ortogonalna_Korelacija_Pravokotni_primer2.m
- m-datoteke-Pogl_3_3\ID_Korelacijska_analiza_Primer1.m
- m-datoteke-Pogl_3_3\Pristranskost_ocene_KK_F.m
- m-datoteke-Pogl_3_3\Pristranskost_ocene_KK_F_sumni_sig.m
- m-datoteke-Pogl_3_3\ID_Korelacijska_analiza_Primer2.m
- m-datoteke-Pogl_3_3\ID_Korelacijska_analiza_Primer3.m
- m-datoteke-Pogl_3_3\ID_Korelacijska_analiza_Primer4.m
- m-datoteke-Pogl_3_3\ID_Korelacijska_analiza_Primer5.m
- m-datoteke-Pogl_3_3\ID_Korelacijska_analiza_Primer6.m
- m-datoteke-Pogl_3_4\ID_Spektralna_analiza_Primer_1.m
- m-datoteke-Pogl_3_4\ID_Spektralna_analiza_Primer_2.m
- m-datoteke-Pogl_6_1\Identifikabilnost_zaprte_zanke_Neparametricni.m
- m-datoteke-Pogl_6_1\Identifikabilnost_zaprte_zanke_Parametricni.m