Amar Thelidji University - Laghouat

Faculty of Technology

Department of Electrical Engineering

Master's Degree on Automatic Control and Systems

Course Material

Identification Techniques

Author :
Dr. Rahmani Belkacem

Contents

List of Figures

List of Tables

List of Abbreviations

MPC	Model Predictive Control
SISO	Single Input Single Output.
MIMO	Multi Input Multi Output.
ACF	Auto-Correlation Function.
PRBS	Pseudo Random Binary Signal.
LTI	Linear Time Invariant System.
ZOH	Zero Order Hold.
ADC	Analog Digital Converter.
DAC	Digital Analog Converter.
ARX	AutoRegressive eXogenous input
ARMAX	AutoRegressive Moving-Average model with eXogenous inputs.
FIR	Finite Impulse Response.
OE	Output Error Model.
BJ	Box Jenkins.

List of Symbols

$\mathbf{Z}\{.\}$	Z transform
$\mathbf{L}\{.\}$	Laplace transform
$E\{.\}$	Expected value
$R_{ee}()$	Auto-correlation function of $e(k)$
$R_{yu}()$	Inter-correlation function of $y(k)$, $u(k)$
$y(k)$	Measured output at the instant k
$y_p(k)$	Process output at the instant k
$\hat{y}(k)$	Predicted output at the instant k
$n(k)$	Noise signal
$e(k)$	White noise signal
$\varepsilon(k)$	Prediction error
$\phi(k)$	Prediction error
$G(q^{-1})$	Transfer function of the process
q^{-1}	Delay operator
$B(q^{-1})$	Numerator of transfer function
$A(q^{-1})$	Denominator of transfer function
n	Order of the denominator $A(q^{-1})$
m	Order of the Numerator $B(q^{-1})$
d	Pure delay
$\phi(k)$	Regressor vector
$J(k,\theta)$	Cost function
θ	Parameters vector

Chapter 1

Introduction to System Identification

1.1 Introduction

System Identification revolves around the determination of mathematical models based on observed data, making it a field rooted in the principles of data science. It forms the core of all data-driven and measurement-based processes, such as design, control, and monitoring, where models play a pivotal role. The concept of learning from observations is timeless, but the present-day complexities of this discipline find their origins in mathematical and probability theory. This textbook aims to provide a foundational understanding and an overview of special aspect of the data driven modeling, namely parametric identification with a particular focus on linear system.

The analysis of process characteristics and inter-variable relationships is of paramount importance in various aspects of process systems, including prediction, control, monitoring, design, and innovation. A critical step in these analyses is the development of a mathematical representation of the process under study, known as the model.

Two contrasting approaches are generally followed for model development:

1. The theoretical (First-principles) approach, which is based on the fundamental laws of physics.
2. The empirical approach, which is based on the analysis of observations, either from experimental data or operating data.

The empirical approach is a highly practical alternative to the theoretical approach, as most processes are too complex to be understood at a fundamental level. Observations potentially carry a lot of informations that remain otherwise unexploited in a first-principles approach.

The subject of System Identification is concerned with the means and techniques for studying a process system through observed or experimental data, primarily for developing a suitable mathematical description of that system. Figure (1.1) shows a typical practice in System identification. The objective is to develop a model from input-output data. The resulting model is said to be empirical in contrast to being a first-principles model.

The identified model in Figure (1.1) consists of two components:

1. A mathematical description of the cause-and-effect relationships, usually known as the deterministic model.

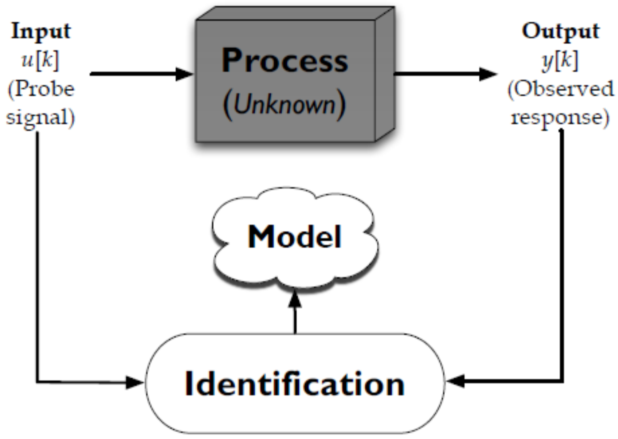

FIGURE 1.1: Identification is mathematical representation of the process [2].

2. A statistical-plus-mathematical description of uncertainties, known as the stochastic model. The stochastic model accounts for observation errors, process uncertainties, and modeling errors.

What factors motivate the need for identification? The two primary motivating factors are:

1. The need for models in process analysis and automation.
2. The practical limitations of the first-principles approach in developing models.

The art of model identification have many application in the field of engineering and in control system in particular. In the following section we introduce a brief introduction for these applications.

1.2 System Identification and Applications:

Before we delve into the main applications for which identification is useful, we firstly define system identification

Definition

System identification is a process in which the mathematical models of dynamic systems are determined or estimated based on observed input and output data.

Once a system is identified, meaning its mathematical models and parameters are determined through the process of system identification, these models find applications across various domains. The applications of system identification are diverse and crucial in understanding and managing complex systems. Figure (1.2) depicts some of the applications of the models to four major branches of process systems engineering, namely, design, estimation (prediction), control, and monitoring.

In all these uses, we create models to simulate and predict how things work. Simulations use computers to save time and money compared to actual experiments, which can be slow, costly,

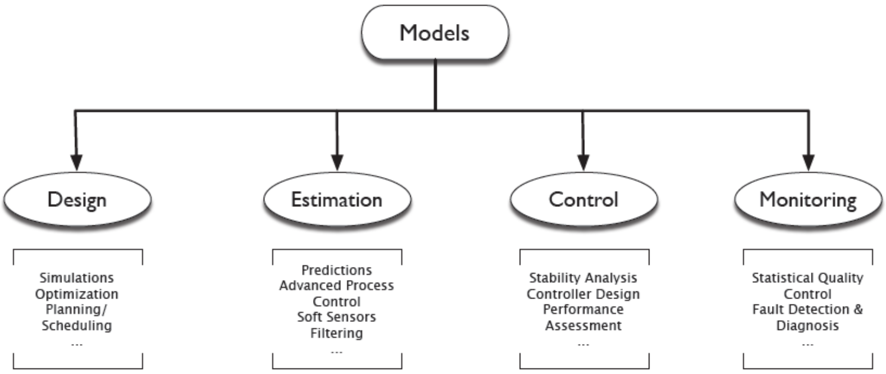

FIGURE 1.2: Applications of models in process systems engineering [2]

and risky. Predictions, on the other hand, are real-time guesses where the model tells us what might happen in the near future during regular conditions. Predictions are handy for designing, controlling, spotting problems, trying out new ideas, and more. But, predictions don't need to be as exact as simulations, where we demand more accuracy.

1.2.1 Modeling of systems

Modeling of system is defined as

Definition

Based on set of assumptions, mathematical modeling is a method of representing real-world situations or systems using mathematical structures, equations, or formulas.

Generally, when we think about modeling, we talk about two types of modeling strategies

- **First-principle modeling**: First-principle modeling involves building mathematical models rooted in fundamental scientific principles to describe and predict the behavior of a system (white boxe modeling).
- **Data driven modeling**: Data-driven modeling involves constructing mathematical models based on empirical data using statistical and machine learning methods.

Data-driven modeling is considered as an alternative to first-principles modeling, especially for highly complex systems or those where obtaining a model is challenging due to the intricate nature of the system. Two common types of models are prevalent in the field of system identification:

- **Gray box model:** This model incorporates some prior information about the system, such as a known model structure (first modeling principle) or an understanding derived from exploring the step response of the system.
- **Black box model:** In this model, no prior information is available. Both the structure and parameters need to be identified through data-driven approaches.

1.2.2 Controller design

The initial modeling principle comes with limitations, particularly in its ability to consistently capture high-frequency resonance modes. To address this, an identification experiment that specifically stimulates these modes can be conducted, leading to the derivation of an improved model. This enhanced model can subsequently be employed in the design of a controller with a broader bandwidth.

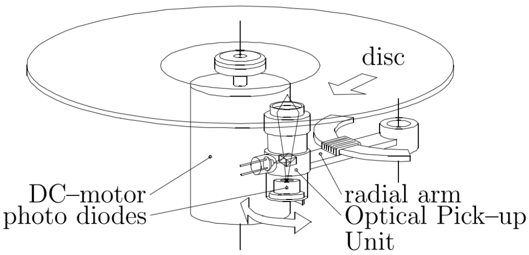

FIGURE 1.3: Guiding mechanism which positions the laser on the correct CD track using a mechanical arm.

Figure (1.3) presents an example of such application where the first principle modeling does not capture all reseneant modes of the guiding mechanism of the laser in the joggobale cd player.

1.2.3 Predictions

System identification is also very useful in making prediction about the behavior of specific phenomena

- **Predictive controllers:** Operate based on dynamic models of the process under control, subject to a predefined set of constraints.
- **In the case of MPC controllers:** While these controllers may relay on physical models at times, they commonly employ ARX (AutoRegressive with eXogenous input) and FIR (Finite Impulse Response) models.
- **Prediction process:** Utilizes past measurements of inputs and outputs, with a predictor constructed using the identified model, to generate forecasts for a specified number of steps ahead.
- **Applications in forecasting:** Extend to diverse scenarios, including predicting floods, assessing avalanche risks, estimating future energy consumption, forecasting solar radiation, and more.

1.2.4 Fault Detection

To effectively monitor a process, we generally follow these key steps:

- **Identify and validate a model:** Develop and validate a model for the process under normal conditions using data collected during regular operations.

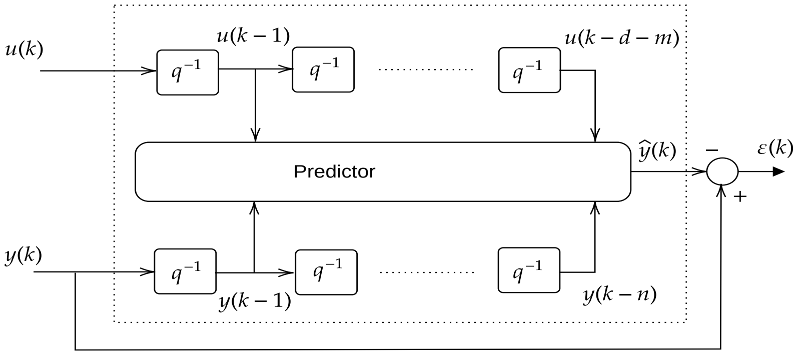

FIGURE 1.4: Fault detection

- **Implement predictor and calculate residuals:** Deploy the predictor associated with the identified model and compute the residuals.
- **Monitor residuals for changes:** Keep track of various characteristics such as mean, variance, whiteness, etc., in the residuals to detect any deviations or defects.

1.3 System Identification Notions

1.3.1 Dynamical system

Definition

A system is an object in which variables of different natures interact and produce observable signals.

- The observable signals of interest are generally called **outputs**.
- Signals that can be manipulated by the user are called **inputs**.
- Others are called **disturbances**:

1.3.2 Identification procedure

Constructing a model from data involves managing four fundamental components:

- **Data:** The data utilized is collected through a specifically designed experiment. The goal of experiment design is to make the data as informative as possible within the given constraints.
- **A set of candidate models:** The crucial and challenging decision in the identification process. This choice relies on a combination of a priori information and various formal properties of the models.
- **An evaluation rule:** This rule determines how candidate models are assessed using the available data. Model quality evaluation typically revolves around the model's capacity to accurately reproduce measured data.

- **Model validation:** The process of testing whether the identified model is deemed "good enough" for its intended purpose.

1.3.3 Design of Experiments

To ensure optimal data collection for the system, a good design of the experiment is essential. This encompasses various aspects, particularly in input design (including the selection of the input, its amplitude, experiment duration, etc.). Some key considerations for linear systems are outlined below:

- The experiment should be conducted under conditions as closely resembling those in which the model will be utilized.
- For Single Input Single Output (SISO) systems: The input should stimulate all relevant frequencies. A suitable choice is a Pseudo Random Binary Sequence (PRBS) signal.
- The sampling frequency must be appropriately chosen.
- For Multi-variable systems (MIMO): The input signals should be independent of each other.

The general identification procedure is illustrated in the following diagram

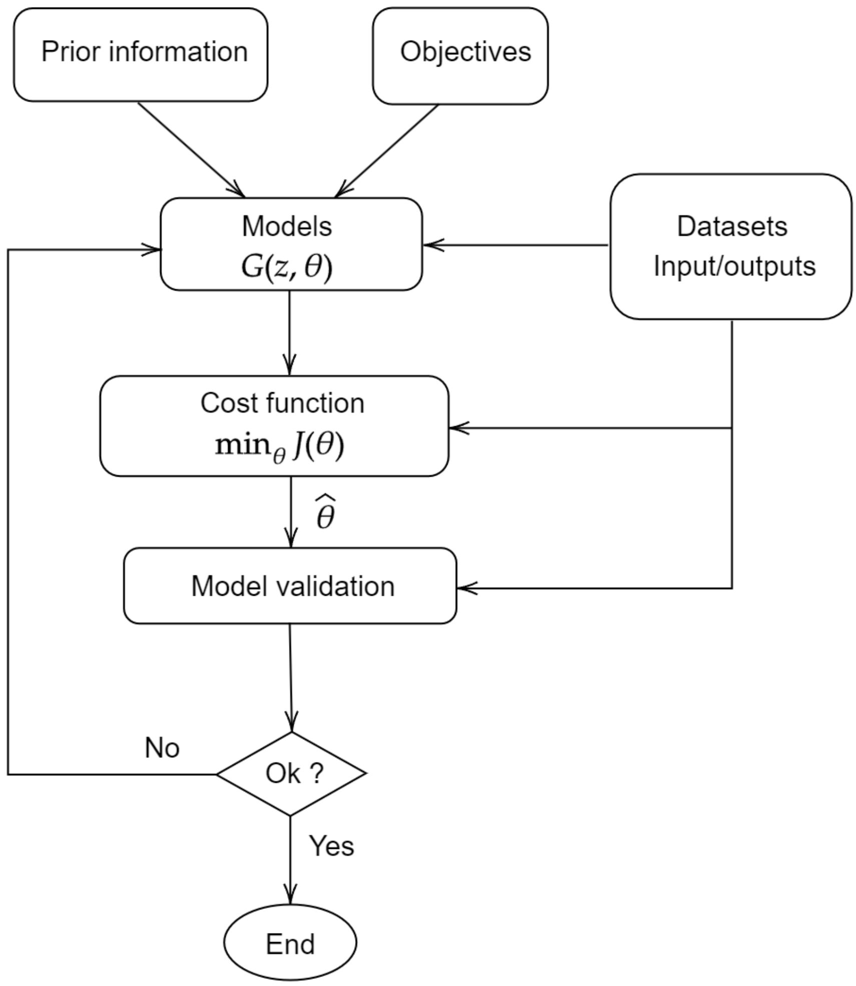

FIGURE 1.5: System identification procedure

Chapter 2

Theoretical Notions

2.1 Representation of a discrete system

2.1.1 Time representation

Consider a discrete, single-variable dynamic system with a sampling period denoted as T, and let $y(k) \equiv y(kT)$ represent its output. For a Linear Time Invariant (**LTI**) system, the output is expressed through the convolution product given by:

$$\begin{aligned} y(k) &= \sum_{j=0}^{k} g(k-j)u(j) = \sum_{j=0}^{k} g(j)u(k-j) \qquad k = 0, 1, 2, \ldots \\ y(k) &= g(k) * u(k) = u(k) * g(k) \end{aligned} \tag{2.1}$$

Here, $y(k)$ is the output at time step k, $u(j)$ represents the input, and $g(j)$ is the system's impulse response. This convolution equation captures the relationship for an LTI system, emphasizing its dependency on the convolution of the input and the system's impulse response.

2.1.2 Transfer Function Representation

A representation based on the transform of temporal signals (Laplace transform, z transform) is often useful. We define:

$$G(s) = \frac{Y(s)}{U(s)} = \mathcal{L}[g(t)] \equiv \int_0^\infty g(t)e^{-st}\, dt \tag{2.2}$$

Or in discrete form:

$$G(z) = \frac{Y(z)}{U(z)} = \mathcal{Z}[g(k)] \equiv \sum_{k=0}^{\infty} g(k)z^{-k} \tag{2.3}$$

This representation is only valid for Linear Time Invariant (LTI) systems.

Note: Discrete time to continuous time conversion is used when $G(z)$ is available by identification, and we need $G(s)$ for continuous time design or for the identification of physical parameters.

2.1.3 Evaluation of $G(z)$ from $G(s)$

For a given $G(s)$, the function $G(z)$ depends on the input $u(t)$. In general, the transfer function is given by:

$$G(z) = \frac{Y(z)}{U(z)} = \frac{\mathcal{Z}\{y(k)\}}{\mathcal{Z}\{u(k)\}} = \frac{\mathcal{Z}\left\{\mathcal{L}^{-1}[G(s)U(s)]|_{kT}\right\}}{\mathcal{Z}\left\{\mathcal{L}^{-1}[U(s)]|_{kT}\right\}} \tag{2.4}$$

Evaluation of $G(z)$ considering the input $u(t)$ generated by a zero-order hold ZOH

This approach is crucial in control systems as it corresponds to the case where the process is surrounded by converters DAC (with a zero-order hold ZOH) and ADC, as shown in the following figure. The transfer function can be written as:

$$G(z) = \frac{\mathcal{Z}\left\{\mathcal{L}^{-1}[G(s)/s]|_{kT}\right\}}{\frac{1}{1-z^{-1}}} = \left(1 - z^{-1}\right)\mathcal{Z}\left\{\mathcal{L}^{-1}[G(s)/s]\Big|_{kT}\right\} \tag{2.5}$$

a

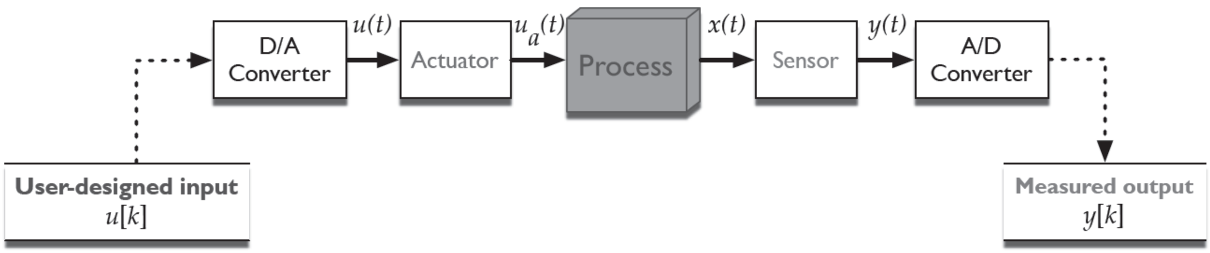

FIGURE 2.1: A typical sampled-data system [2].

Example 2.1.1 *Let the analog transfer function* $G(s) = \frac{1}{s+1}$*. Let us evaluate* $G(z)$ *for* $T = 0.2$.

$$G(z) = \left(1 - z^{-1}\right)\mathcal{Z}\left\{\mathcal{L}^{-1}\left[\frac{1}{s(s+1)}\right]\Bigg|_{kT}\right\}$$

$$G(z) = \left(1 - z^{-1}\right)\mathcal{Z}\left\{\mathcal{L}^{-1}\left[\frac{1}{s} - \frac{1}{(s+1)}\right]\Bigg|_{kT}\right\} = \left(1 - z^{-1}\right)\left[\frac{1}{1-z^{-1}} - \frac{1}{1-e^{-T}z^{-1}}\right]$$

$$G(z) = 1 - \frac{1-z^{-1}}{1-e^{-T}z^{-1}} = \frac{\left(1-e^{-T}\right)z^{-1}}{1-e^{-T}z^{-1}} = \frac{0,18z^{-1}}{1-0.82z^{-1}}$$

Evaluation of $G(\mathbf{z})$ by finite differences:

In the context of the differential equation corresponding to $G(s)$, we can employ a finite difference method for sampling.

Euler Backward Integration:

The Euler backward integration method, a numerical approach for solving ordinary differential equations (ODEs), approximates continuous derivatives and integrals discretely. For the transfer function:

$$G(s) = \frac{Y(s)}{U(s)} = \frac{1}{s} \tag{2.6}$$

or, in the time domain:

$$\dot{y}(t) = u(t) \quad \text{and} \quad y(0) = 0 \tag{2.7}$$

The trajectory $y(t)$ can be approximated using backward rectangular integration:

$$y(k) = y(k-1) + Tu(k) \tag{2.8}$$

This leads to the discrete transfer function:

$$G(z) = \frac{Tz}{z-1} \tag{2.9}$$

From this, we derive the approximation:

$$\frac{1}{s} \approx \frac{Tz}{z-1} \tag{2.10}$$

or equivalently:

$$s \approx \frac{1}{T}\frac{z-1}{z} \tag{2.11}$$

Substituting this back into $G(s)$, we get:

$$G(z) = G(s)|_{s=\frac{1}{T}\frac{z-1}{z}} \tag{2.12}$$

Tustin integration (trapezoidal):

This method represents another approach of approximating derivatives and integrals using trapezoidal integration. The trajectory $y(t)$ is approximated as follows:

$$y(k) = y(k-1) + T\frac{u(k) + u(k-1)}{2} \tag{2.13}$$

This leads to the discrete transfer function:

$$Y(z)\left[1 - z^{-1}\right] = \frac{T}{2}U(z)\left(1 + z^{-1}\right)$$
$$G(z) = \frac{Y(z)}{U(z)} = \frac{T}{2}\left(\frac{1+z^{-1}}{1-z^{-1}}\right) = \frac{T(z+1)}{2(z-1)}$$

And the approximation for $\frac{1}{s}$ is given by:

$$\frac{1}{s} \approx \frac{T(z+1)}{2(z-1)} \quad \text{or} \quad s \approx \frac{2}{T}\frac{z-1}{z+1}$$

Therefore:

$$G(z) = G(s)|_{s=\frac{2}{T}\frac{z-1}{z+1}} \tag{2.14}$$

Example 2.1.2 *Given the transfer function $G(z) = \frac{1}{s+1}$ and $T = 0.2$ s, calculate $G(z)$ using backward integration and Tustin methods:*

$$G(z) = \frac{1}{s+1}\bigg|_{s=\frac{1}{T}\frac{z-1}{z}} = \frac{1}{\frac{1}{T}\frac{z-1}{z}+1} = \frac{Tz}{(1+T)z-1} = \frac{0.2}{1.2-z^{-1}}$$

$$G(z) = \frac{1}{s+1}\bigg|_{s=\frac{2}{T}\frac{z-1}{z+1}} = \frac{1}{\frac{2}{T}\frac{z-1}{z+1}+1} = \frac{T(z+1)}{(2+T)z-(2-T)} = \frac{0.1(1+z^{-1})}{1.1-0.9z^{-1}}$$

2.1.4 Evaluation of $G(s)$ from $G(z)$

The reverse process, i.e., the evaluation of $G(s)$ from $G(z)$, is not unique. However, we can reverse the previous methods to obtain the continuous transfer functions.

Evaluation of $G(s)$ using the inverse ZOH method:

The continuous transfer function can be obtained using the inverse ZOH method:

$$G(s) = s\mathcal{L}\left\{\mathcal{Z}\left[\frac{G(z)}{1-z^{-1}}\right]\right\} \tag{2.15}$$

This involves applying Laplace and inverse Z-transform to $G(z)$ with an additional factor of s.

The inverse Euler Backward Integration method:

The continuous transfer function using the inverse Euler Backward Integration method is given by:

$$G(s) = G(z)|_{z=\frac{1}{1-Ts}} \tag{2.16}$$

This involves substituting z with $\frac{1}{1-Ts}$ in the discrete transfer function.

The inverse Tustin method:

The continuous transfer function using the inverse Tustin method is given by:

$$G(s) = G(z)|_{z=\frac{1+\frac{T}{2}s}{1-\frac{T}{2}s}} \tag{2.17}$$

This involves substituting z with $\frac{1+\frac{T}{2}s}{1-\frac{T}{2}s}$ in the discrete transfer function.

2.2 Notions on Signal Analysis

In the realm of signal analysis, correlation functions play a crucial role in understanding the relationships between different types of signals. Two distinct signal categories are often considered, characterized by their energy content and average power.

2.2.1 Correlation Functions for Deterministic Signals

Let's delve into the definitions, starting with transient signals possessing finite energy, exemplified by the signal $u(k)$:

$$0 \leq \sum_{k=-\infty}^{\infty} u^2(k) < \infty \tag{2.18}$$

On the other hand, periodic or random signals exhibit finite average power:

$$0 \leq \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} u^2(k) < \infty \tag{2.19}$$

Correlation functions serve as measures of **similarity in shape and position** between two signals, notably the auto-correlation function (ACF) $R_{uu}(h)$ and the inter-correlation $R_{uy}(h)$, which describe the interdependence of $u(k)$ and $y(k+h)$.

In practical terms, when dealing with a **finite number of data** from $u(k)$ and $y(k)$ ($k = 0, 1, \ldots, (N-1)$), approximate evaluations for these correlation functions are employed.

Finite Energy Signals

For real signals with finite energy and N values of the translation parameter h, we obtain:

Auto-correlation

$$\bar{R}_{uu}(h) = \sum_{k=0}^{N-1} u(k)u(k+h) \quad h = 0, 1, \ldots, N-1 \tag{2.20}$$

Inter-correlation

$$\begin{aligned}\bar{R}_{uy}(h) &= \sum_{k=0}^{N-1} u(k)y(k+h) = \sum_{k=0}^{N-1} u(k-h)y(k) \quad h = 0, 1, \ldots, N-1 \\ \bar{R}_{yu}(h) &\equiv \sum_{k=0}^{N-1} y(k)u(k+h) = \sum_{k=0}^{N-1} y(k-h)u(k) \quad h = 0, 1, \ldots, N-1\end{aligned} \tag{2.21}$$

2.2.2 Signals with Finite Average Power

For signals with finite average power and a finite number of data $2N-1$, we utilize the following definitions.

Auto-correlation

$$\bar{R}_{uu}(h) = \frac{1}{N} \sum_{k=0}^{N-1} u(k)u(k+h) = \frac{1}{N} \sum_{k=0}^{N-1} u(k-h)u(k) \quad h = 0, 1, \ldots, N-1 \tag{2.22}$$

Inter-correlation

$$\begin{aligned} \bar{R}_{uy}(h) &\equiv \frac{1}{N}\sum_{k=0}^{N-1} u(k)y(k+h) = \frac{1}{N}\sum_{k=0}^{N-1} u(k-h)y(k) \\ \bar{R}_{yu}(h) &\equiv \frac{1}{N}\sum_{k=0}^{N-1} y(k)u(k+h) = \frac{1}{N}\sum_{k=0}^{N-1} y(k-h)u(k) \end{aligned} \tag{2.23}$$

When $u(k)$ and $y(k)$ are periodic with period N, the above relationships present the exact values of the auto-correlation and cross-correlation functions, also periodic with the same period.

Properties

We demonstrate the following properties using the definitions of correlation functions:

- The autocorrelation function of a real signal is an even function:

$$\bar{R}_{uu}(h) = R_{uu}(-h) \tag{2.24}$$

- Considering the relative position of the signals u and y, a translation h of u is equivalent to a translation $(-h)$ of y, hence the relation:

$$\bar{R}_{yu}(h) = R_{uy}(-h) \tag{2.25}$$

On the other hand, for all h:

$$\begin{aligned} \left|\bar{R}_{uu}(h)\right| &\leq R_{uu}(0) \\ \left|\bar{R}_{uy}(h)\right| &\leq \sqrt{R_{uu}(0)R_{yy}(0)} \leq \frac{1}{2}\left[R_{uu}(0) + R_{yy}(0)\right] \end{aligned} \tag{2.26}$$

2.3 Random signals and Processes

This section is devoted to the concepts of discrete-time random signals and processes, particularly the notions of realization, stationarity and ergodicity.

Definition

A (discrete-time) **random signal** is a function $x(k)$ whose characteristics cannot be accurately described by any existing mathematical function. At each point, it is characterized by a it probabilistic characteristics

An alternative definition

Definition

A (discrete-time) **random signal**

$$x = \{x(1), x(2), \ldots, x(k), x(k+1), \cdots, x(N)\} \tag{2.27}$$

is an index-ordered sequence of random variables in time. Its characteristics can only be described by probabilistic laws and not merely by mathematical models.

Definition

A single record is only one of the numerous possible sets of observations. For this reason, a finite length record of a random process is said to be a **realization** of the process.

As in the case of random variables, all possible realizations of a random signal constitute the **ensemble** of that signal or process.

Definition

The process (or a subsystem) that generates a random signal is termed as the **random process** or **stochastic** process.

Definition

A **deterministic process** on the other hand generates signals whose values are accurately known (no uncertainty) and accurately predictable (completely understood).

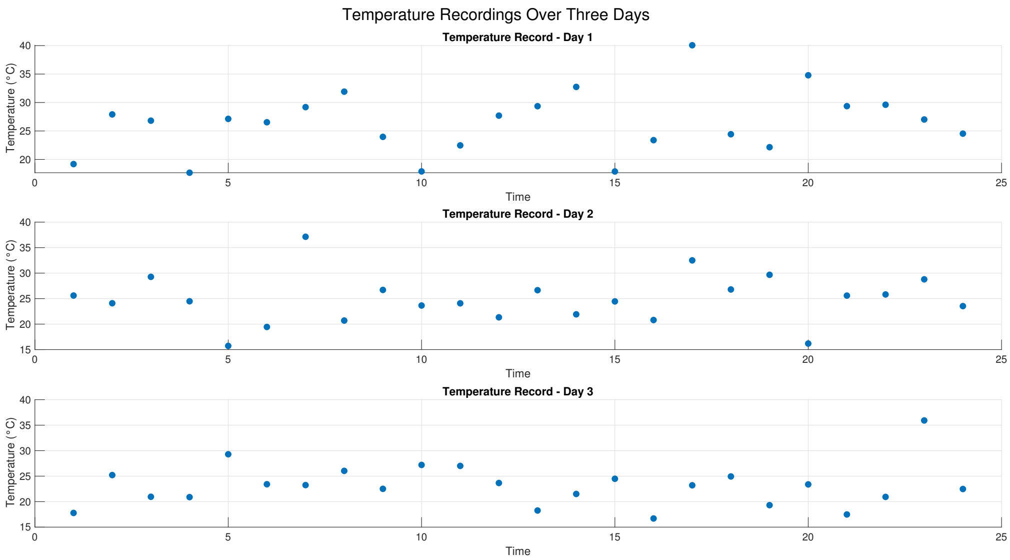

FIGURE 2.2: Three-day recording of a temperature variable

As an example, consider the output of a process recorded for hours and several days (Figure (2.2)). The evolution of the output can be described by a deterministic function $y(k)$, but this function will be different from day to day. At a given time, we measure the value of the output every day. Each of these functions is called a **realization** of the temperature measurement. Although, the temperature is continuous variable, Figure (2.2) presents its discrete counterpart. For all the values measured every day at the same time, we can define a statistic characterized by **an average value** and **a standard deviation.**.

Random signals are seen as collections of random variables, where each variable represents the value of the signal at a specific point in time. The concept of ergodicity in the context of random signals is crucial for understanding the statistical properties of these signals. In a broad sense, an ergodic process possesses statistical properties that remain constant over time, and statistical averages computed from a single realization of the process converge to the ensemble averages as the observation time increases.

For a random signal to be ergodic, the statistical behavior of one realization should be representative of the entire ensemble of possible realizations. In simpler terms, analyzing one instance of the signal should yield insights into the overall statistical characteristics of the entire signal. This property simplifies the analysis of random signals, as it allows us to make predictions and draw conclusions about the signal's behavior based on observations from a single instance.

Note 1 *The characteristics of random signal used in practice are the average value (Or Expected value) and variance of the signal. The definition of theses quantities can be very complicated to compute in practice. Practionors in system identification use a modified approximations for random signals with finite set of points. These approximations are known as* ***sample mean*** *and* ***sample variance****.*

If the statistical properties computed from a single realization accurately reflect the overall statistical behavior of a stochastic process, the process is considered to be **ergodic**. In other words, the characteristics observed in one instance of the process are representative of the ensemble's statistical properties. For such signals we define the concept of sample means and variances.

Definition

The sample mean above is only an estimator of the theoretical mean of the signal. Assuming that $x(k)$ is **ergodic**, the sample mean or the expected value of the random signal $x(k)$ is defined by:

$$\mu_x = E\{x(k)\} \cong \frac{1}{N} \sum_{k=0}^{N-1} x(k) \tag{2.28}$$

Definition

The sample mean is used to calculate the sample variance σ_x^2,which can be estimated as follows

$$\sigma_x^2 = E\left\{(x(k) - \mu_x)^2\right\} \cong \frac{1}{N}\sum_{k=0}^{N-1}(x(k) - \mu_x)^2 = E\left\{x(k)^2\right\} - \mu_x^2 \tag{2.29}$$

Or the standard deviation is calculated as

$$\sigma_x = \sqrt{\sigma_x^2} \tag{2.30}$$

An illustration of the statistical characteristics of a random signal is presented in Figure (2.3), where lines representing the mean and standard deviation are shown for each random signal

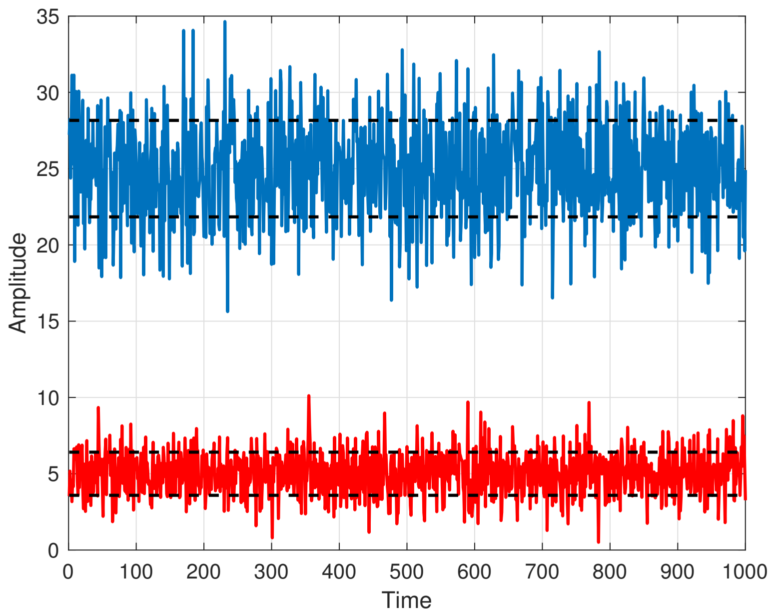

FIGURE 2.3: Illustration of variance and standard deviation, horizontal lines indicate voltage levels that are one standard deviation above and below the mean

Definition

If we have a vector of random variables, we can generalize the notion of variance towards the notion of **covariance** matrix. For a vector of random variables,

$$\boldsymbol{X} = \begin{bmatrix} x_1(k) & x_2(k) & \cdots & x_q(k) \end{bmatrix}^T \tag{2.31}$$

The **covariance** matrix is given by

$$\mathrm{cov}(\boldsymbol{X}) = \mathrm{E}\left[(\boldsymbol{X} - \mu_{\boldsymbol{X}})(\boldsymbol{X} - \mu_{\boldsymbol{X}})^T\right]$$
$$\mathrm{cov}(\boldsymbol{X}) = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_q} \\ \sigma_{x_2x_1} & \sigma_{x_2}^2 & \cdots & \sigma_{x_2x_q} \\ \vdots & \cdots & \cdots & \vdots \\ \sigma_{x_qx_1} & \sigma_{x_qx_2} & \cdots & \sigma_{x_q}^2 \end{bmatrix} = E\left[\boldsymbol{X}\boldsymbol{X}^T\right] - \mu_{\boldsymbol{X}}\mu_{\boldsymbol{X}}^T \tag{2.32}$$

The diagonal of $\mathrm{cov}(\boldsymbol{X})$ contains the variance of individual random variables while the off-diagonals contain the covariance between a pair of random variables.

For a random signal with finite length, calculating the auto and cross-correlation functions can be challenging. As a result, we often rely on estimates of these functions known as sample correlation functions

Definition

We define biased estimation of the autocorrelation function of a random signal $x(k)$

$$R_{xx}(h) = E\{x(k)x(k+h)\} \cong \frac{1}{N}\sum_{k=0}^{N-h-1} x(k)x(k+h) \quad h = 0, 1, \ldots, N-1 \tag{2.33}$$

The estimation of the autocorrelation function in this case is said to be biased, because the summation is done over $N - h$ terms but the division over N.

Definition

The **unbiased** estimate of autocorrelation function is given as:

$$R_{xx}(h) = E\{x(k)x(k+h)\} \cong \frac{1}{N-h}\sum_{k=0}^{N-h-1} x(k)x(k+h) \quad h = 0, 1, \ldots, N-1 \tag{2.34}$$

This estimate is **unbiased** but has a very large variance if $(N - h)$ is small, because in this case the estimate of the autocorrelation function is based on a small number of data.

Definition

A biased estimate of the cross-correlation function for the ergodic signals $x(k)$ and $y(k)$ is given by:

$$R_{xy}(h) = E\{x(k)y(k+h)\} \cong \frac{1}{N}\sum_{k=0}^{N-h-1} x(k)y(k+h) \quad h = 0, 1, \ldots, N-1 \tag{2.35}$$

And an unbiased estimate by:

$$R_{xy}(h) = E\{x(k)y(k+h)\} \cong \frac{1}{N-h}\sum_{k=0}^{N-h-1} x(k)y(k+h) \quad h = 0, 1, \ldots, N-1 \tag{2.36}$$

If $x(k)$ and $y(k)$ are independent, we have:

$$R_{xy}(h) = E\{x(k)\}E\{y(k+h)\} \qquad h = 0, 1, \ldots, N-1 \tag{2.37}$$

And if $x(k)$ and $y(k)$ are independent and have a zero mean value, we will have:

$$R_{xy}(h) = E\{x(k)y(k+h)\} = 0 \quad \forall\ h \tag{2.38}$$

Note 2 *In terms of notation, we use Rxx to denote the sample estimate of the autocorrelation function, and $\bar{R}xx$ for the average autocorrelation. The same convention applies to cross-correlations.*

2.3.1 Discreet white noise

Most random disturbances can be described as discrete white noise applied to a filter.

Definition

White noise is a realization of a random process in which the power spectral density $\phi_{ee}(\omega)$ is the same for all frequencies.

Definition

white noise is a stochastic process where the autocorrelation function is an impulse function at zero lag and is zero for all other lags. Mathematically, the autocorrelation function $R_{ee}(h)$ of white noise is given by:

$$R_{ee}(h) = \begin{cases} \sigma_e^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases} \tag{2.39}$$

Here, σ_e^2 represents the variance of the white noise.

This definition implies that the values of the white noise at different time points are uncorrelated (except for the case when the time points are the same, where the autocorrelation is equal to the variance). Discrete white noise and its auto-correlation and spectral density are illustrated in Figure (2.4). We subsequently consider white noise with zero mean value $e(k)$ as a signal with the following properties

$$E\{e(k)\} = 0 \quad \forall k$$
$$R_{ee}(h) = E\{e(k)e(k+h)\} = \begin{cases} \sigma_e^2 & \text{for } h = 0 \\ 0 & \text{else} \end{cases} \tag{2.40}$$

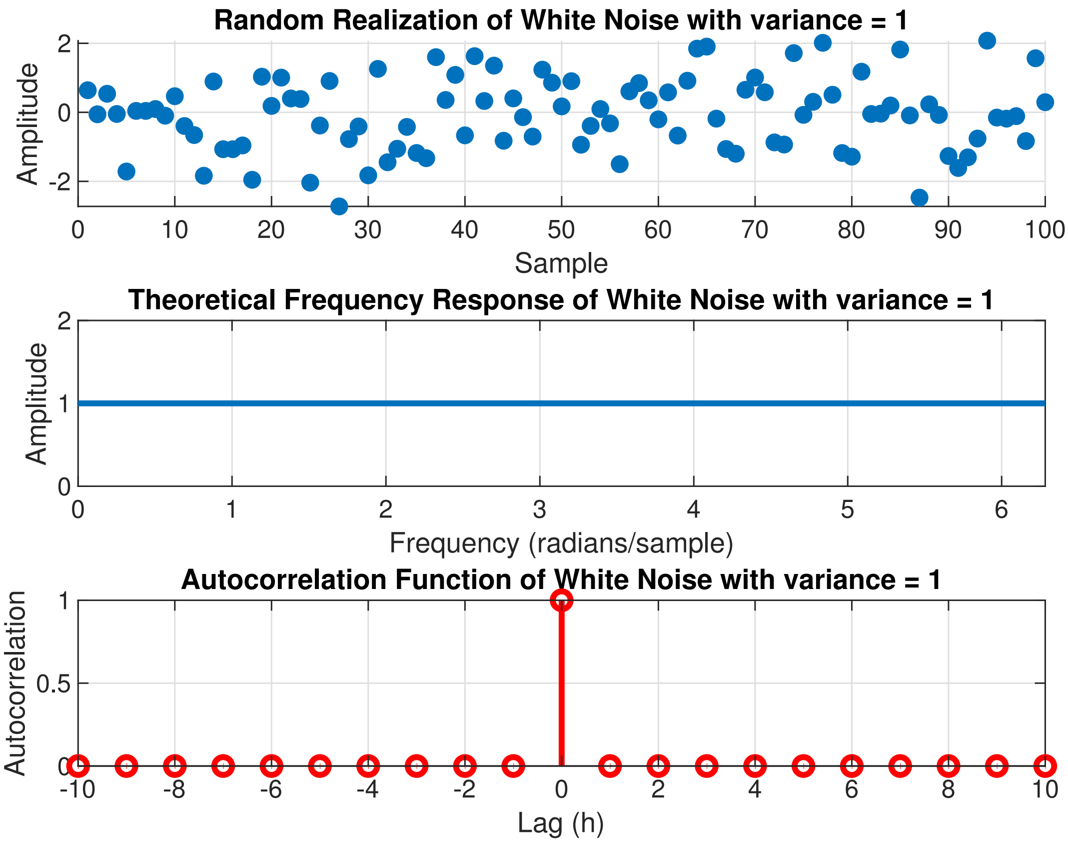

FIGURE 2.4: Discrete white noise $e(k)$ with its auto-correlation $R_{ee}(h)$ and its spectral density $\varphi_{ee}(\Omega)$.

These properties do not fully define the signal because $e(k)$ can follow different distributions (uniform, normal or Gaussian, etc.). In practice, we often use the Gaussian distribution $\mathcal{N}(0, \sigma_e^2)$. Figure (2.5) illustrates the biased and unbiased estimate of the auto-correlation function of a discrete white noise realization with $N = 1000$ data.

2.3.2 ACF of a Moving Average (MA) process

Consider an MA (moving average) process where the noise is modeled as:

$$n(k) = \left(1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}\right) e(k) = C\left(q^{-1}\right) e(k) \tag{2.41}$$

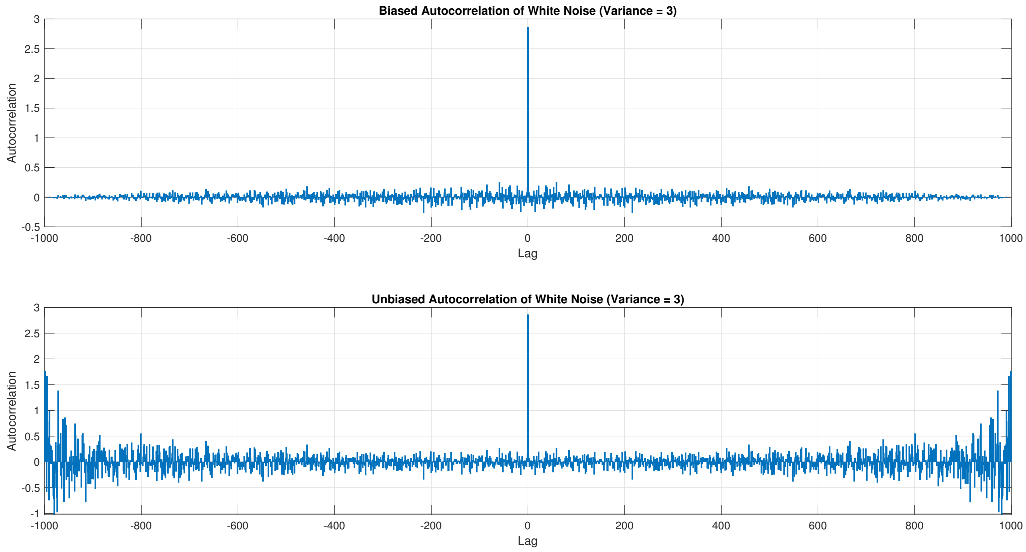

FIGURE 2.5: The estimation of the auto-correlation function of a realization of discrete white noise with 1000 data (a) biased and (b) unbiased

Taking for example $n_c = 1$, The average of $n(k)$ is given by

$$\begin{aligned} E\{n(k)\} &= E\left\{\left(1 + c_1 q^{-1}\right) e(k)\right\} = E\left\{e(k) + c_1 e(k-1)\right\} \\ &= E\{e(k)\} + c_1 E\{e(k-1)\} = 0 \end{aligned}$$

The variance of $n(k)$

$$E\{n(k)n(k)\} = E\left\{[e(k) + c_1 e(k-1)]^2\right\} \qquad = E\left\{e^2(k)\right\} + c_1^2 E\left\{e^2(k-1)\right\} + 2c_1 E\{e(k)e(k-1)\}$$

The third term is zero due to the non-correlation of white noise, thus;

$$R_{nn}(0) = E\{n(k)n(k)\} = \left(1 + c_1^2\right)\sigma_e^2 \tag{2.42}$$

To calculate $R_{nn}(h)$ for $h \neq 0$ we proceed as follows. For $h = 1$:

$$\begin{aligned} n(k-1) =& e(k-1) + c_1 e(k-2) \\ R_{nn}(1) =& E\{n(k)n(k-1)\} = E\left\{[e(k) + c_1 e(k-1)]\,[e(k-1) + c_1 e(k-2)]\right\} \\ =& E\{e(k)e(k-1)\} + c_1 E\{e(k)e(k-2)\} + c_1 E\{e(k-1)e(k-1)\} \\ &+ c_1^2 E\{e(k-1)e(k-2)\} \\ =& c_1 \sigma_e^2 \end{aligned}$$

Then we obtain

$$R_{nn}[h] = \begin{cases} (1+c_1^2)\,\sigma_e^2 & h = 0 \\ c_1\sigma_e^2 & h = \pm 1 \\ 0 & h = \pm 2, \pm 3, \cdots \end{cases}$$

The Autocorrelation function of the moving avarage is illustrated in the following Figure

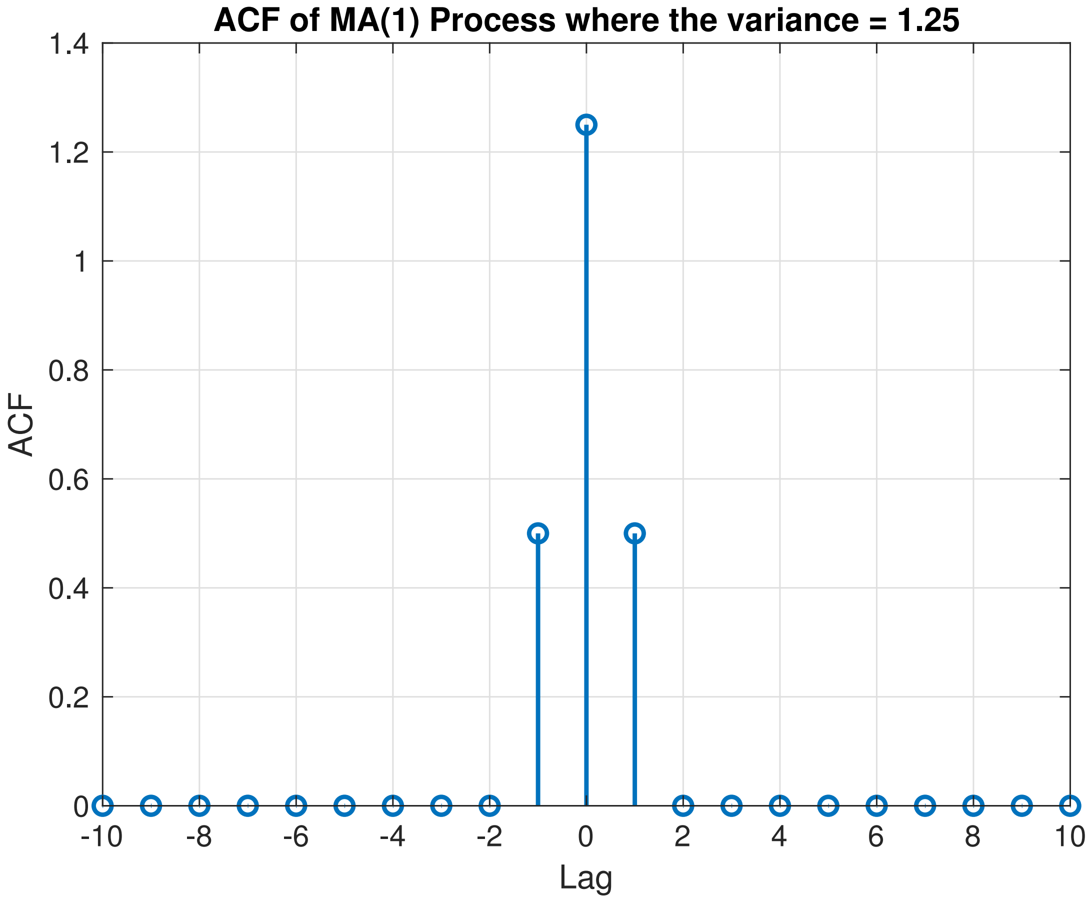

FIGURE 2.6: Autocorrelation function for a first order MA process

2.4 Solved Exercises

Exercise 2.4.1 *The analog system* $G(s) = \frac{2e^{-s}}{5s+1}$ *needs to be controlled by a microprocessor.*

1. *Calculate* $G(z)$*.*
2. *Determine the poles of the discrete system.*
3. *Calculate the response of the discrete system to a unit impulse.*

Solution:
Important: *A sampling period must be specified for discretization. Let's choose* $T = 1$ *second for convenience.*

1. ***Calculate*** $G(z)$

Given the continuous-time system:

$$G(s) = \frac{2e^{-s}}{5s+1}$$

With sampling period $T = 1$ second. Using Zero-Order Hold (ZOH) discretization:

$$G(z) = (1 - z^{-1})\mathcal{Z}\left\{\mathcal{L}^{-1}\left[\frac{G(s)}{s}\right]_{t=kT}\right\}$$

Step 1: *Compute without the delay:*

$$\frac{2}{s(5s+1)} = \frac{A}{s} + \frac{B}{5s+1}$$

Solving:

$$2 = A(5s+1) + Bs$$

Comparing coefficients:

$$5A + B = 0, \quad A = 2 \Rightarrow B = -10$$

Thus:

$$\frac{2}{s(5s+1)} = \frac{2}{s} - \frac{10}{5s+1} = \frac{2}{s} - \frac{2}{s+0.2}$$

Inverse Laplace transform:

$$f(t) = 2(1 - e^{-0.2t})$$

Step 2: *Discretize using ZOH with $T = 1$:*

$$F(z) = (1 - z^{-1})\mathcal{Z}\{f(t)\} = (1 - z^{-1})\left[\frac{2}{1 - z^{-1}} - \frac{2}{1 - e^{-0.2}z^{-1}}\right]$$

$$F(z) = 2 - \frac{2(1 - z^{-1})}{1 - e^{-0.2}z^{-1}} = \frac{2(1 - e^{-0.2}z^{-1}) - 2(1 - z^{-1})}{1 - e^{-0.2}z^{-1}}$$

$$F(z) = \frac{2(1 - e^{-0.2})z^{-1}}{1 - e^{-0.2}z^{-1}}$$

Numerically:

$$e^{-0.2} \approx 0.81873, \quad 1 - e^{-0.2} \approx 0.18127$$

So:

$$F(z) = \frac{0.36254z^{-1}}{1 - 0.81873z^{-1}}$$

Step 3: *Include the time delay of 1 second. Since $T = 1$, this corresponds to exactly one sample delay:*

$$G(z) = z^{-1}F(z) = \frac{0.36254z^{-2}}{1 - 0.81873z^{-1}}$$

Multiplying numerator and denominator by z:

$$G(z) = \frac{0.36254z^{-1}}{z - 0.81873} = \frac{0.36254}{z(z - 0.81873)}$$

2. ***Poles of the discrete system***

   From $G(z) = \frac{0.36254}{z(z-0.81873)}$, the poles are:

$$z = 0 \quad \text{and} \quad z = 0.81873$$

   The pole at $z = 0$ corresponds to a delay of one sampling period.

3. ***Response to unit impulse***

   Given:

$$G(z) = \frac{0.36254}{z(z - 0.81873)}$$

   Let $g(k)$ be the impulse response, where $g(k) = \mathcal{Z}^{-1}\{G(z)\}$.

   Method 1: *Direct approach*

$$G(z) = 0.36254 \cdot z^{-1} \cdot \frac{1}{z - 0.81873}$$

   Using the properties:

   - $\mathcal{Z}^{-1}\{z^{-1}F(z)\} = f(k-1)\epsilon(k-1)$ *where $\epsilon(k)$ is the unit step*
   - $\mathcal{Z}^{-1}\left\{\frac{1}{z-a}\right\} = a^{k-1}\epsilon(k-1)$

   Thus:

$$g(k) = 0.36254 \cdot (0.81873)^{k-2} \cdot \epsilon(k-2)$$

   Or explicitly:

$$g(k) = \begin{cases} 0 & \text{for } k = 0, 1 \\ 0.36254 & \text{for } k = 2 \\ 0.36254 \cdot (0.81873)^{k-2} & \text{for } k \geq 3 \end{cases}$$

   Method 2: *Partial fraction expansion*

$$\frac{0.36254}{z(z - 0.81873)} = \frac{A}{z} + \frac{B}{z - 0.81873}$$

$$0.36254 = A(z - 0.81873) + Bz$$

$$0.36254 = (A + B)z - 0.81873A$$

Equating coefficients:

$$A + B = 0 \quad \textit{and} \quad - 0.81873A = 0.36254$$

$$A = -0.44295, \quad B = 0.44295$$

Thus:

$$G(z) = -0.44295 \cdot \frac{1}{z} + 0.44295 \cdot \frac{1}{z - 0.81873}$$

Inverse z-transform:

$$g(k) = -0.44295\delta(k-1) + 0.44295(0.81873)^{k-1}\epsilon(k-1)$$

Exercise 2.4.2 *The design of a proportional-integral (PI) controller for a dynamic system resulted in the following transfer function:*

$$G_R(s) = 0.72\left(1 + \frac{1}{0.06s}\right) = 0.72 + \frac{12}{s}$$

Determine the discrete-time equation required for implementing this controller on a microprocessor.

Solution:

The discrete-time controller will be interfaced with analog-to-digital (A/D) and digital-to-analog (D/A) converters, with the latter utilizing a zero-order-hold element.

- ***Step 1:*** *Specify sampling period. Let's use $T = 0.1$ seconds as a reasonable choice.*
- ***Step 2:*** *Calculate $G_R(z)$ using ZOH method:*

$$\begin{aligned}
G_R(z) &= (1 - z^{-1})\mathcal{Z}\left\{\mathcal{L}^{-1}\left[\frac{G_R(s)}{s}\right]_{t=kT}\right\} \\
&= (1 - z^{-1})\mathcal{Z}\left\{\mathcal{L}^{-1}\left[\frac{0.72}{s} + \frac{12}{s^2}\right]\right\} \\
&= (1 - z^{-1})\left[\frac{0.72}{1 - z^{-1}} + \frac{12Tz^{-1}}{(1 - z^{-1})^2}\right] \\
&= 0.72 + \frac{12Tz^{-1}}{1 - z^{-1}} \\
&= \frac{0.72(1 - z^{-1}) + 12Tz^{-1}}{1 - z^{-1}} \\
&= \frac{0.72 + (12T - 0.72)z^{-1}}{1 - z^{-1}}
\end{aligned}$$

- ***Step 3:*** *With $T = 0.1$:*

$$12T = 12 \times 0.1 = 1.2$$

$$G_R(z) = \frac{0.72 + (1.2 - 0.72)z^{-1}}{1 - z^{-1}} = \frac{0.72 + 0.48z^{-1}}{1 - z^{-1}}$$

- ***Step 4:*** *Write the difference equation:*

$$U(z)(1 - z^{-1}) = E(z)\left[0.72 + 0.48z^{-1}\right]$$

The inverse z-transform results in:

$$u(k) = u(k-1) + 0.72e(k) + 0.48e(k-1)$$

This is the difference equation that needs to be programmed into the microprocessor.

Exercise 2.4.3 ***Statement:*** *Given the signals $u(n)$ and $y(n)$ as follows for 3 sampling points:*

n	0	1	2
$u(n)$	2	1	1
$y(n)$	1	2	2

Calculate the cross-correlation $R_{uy}(h)$.

Solution:

For periodic signals with period $N = 3$, we compute the circular cross-correlation.

- ***For** $h = 0$**:***

$$R_{uy}(0) = \frac{1}{3}\sum_{n=0}^{2} u(n)y(n) = \frac{1}{3}[u(0)y(0)+u(1)y(1)+u(2)y(2)] = \frac{1}{3}[2\cdot1+1\cdot2+1\cdot2] = \frac{1}{3}[2+2+2] = 2$$

- ***For** $h = 1$**:***

$$R_{uy}(1) = \frac{1}{3}\sum_{n=0}^{2} u(n)y(n+1) = \frac{1}{3}[u(0)y(1) + u(1)y(2) + u(2)y(3)]$$

Since y is periodic with period 3: $y(3) = y(0) = 1$

$$R_{uy}(1) = \frac{1}{3}[2 \cdot 2 + 1 \cdot 2 + 1 \cdot 1] = \frac{1}{3}[4 + 2 + 1] = \frac{7}{3}$$

- ***For** $h = 2$**:***

$$R_{uy}(2) = \frac{1}{3}\sum_{n=0}^{2} u(n)y(n+2) = \frac{1}{3}[u(0)y(2) + u(1)y(3) + u(2)y(4)]$$

Since y is periodic: $y(3) = y(0) = 1$ and $y(4) = y(1) = 2$

$$R_{uy}(2) = \frac{1}{3}[2 \cdot 2 + 1 \cdot 1 + 1 \cdot 2] = \frac{1}{3}[4 + 1 + 2] = \frac{7}{3}$$

Thus:

$$R_{uy}(0) = 2, \quad R_{uy}(1) = \frac{7}{3}, \quad R_{uy}(2) = \frac{7}{3}$$

Note: *For real periodic signals,* $R_{uy}(-h) = R_{uy}(N-h)$*, so:*

$$R_{uy}(-1) = R_{uy}(2) = \frac{7}{3}, \quad R_{uy}(-2) = R_{uy}(1) = \frac{7}{3}$$

Chapter 3

Parametric Identification

3.1 Introduction

Parametric identification is a modeling technique employed in system analysis, particularly in dynamic systems, where the goal is to characterize the system's behavior using a predefined set of parameters. Parametric identification involves defining the model's form in advance which employs transfer functions representation, with explicit parameters representing key system characteristics. The challenge lies in estimating and fine-tuning these parameters based on observed input-output data, leading to a more structured and interpretable representation of the underlying system. This chapter considers the parametric representation of dynamic systems based on input-output measurements [1].

- A first element of parametric modeling is to calculate a transfer function from a non-parametric representation.
- In the case of a parametric approach, the structure of the model must be specified a priori, for example in the form of the first order transfer function with pure delay $G(s) = Ke^{-\theta s}/(\tau s+1)$ with the 3 parameters K, τ and θ. This constitutes the main characteristic of parametric models.

In this chapter, the identification of the model parameters will be done on the basis of a criterion on the difference between experimental measurements coming from the process and the computed output of a predictor as shown in Figure (3.1) .

To do this, we must first choose:

- The structure of the model
- The performance criterion,
- The identification algorithm
- The excitation signal u.

The methods presented in this chapter have many advantages:

- Better precision,
- Monitoring of model parameters in real time allowing, if necessary, adjustment of the regulator during process operation

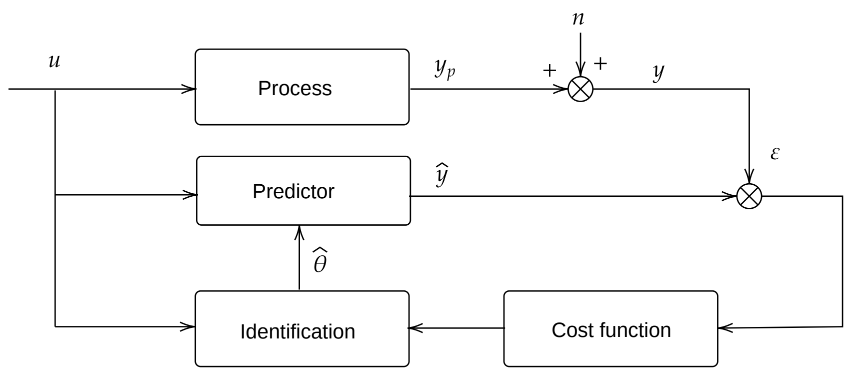

FIGURE 3.1: General scheme of identification: u input (adjustable); y_p, process output (not measurable); n noise (unknown) ; y noisy output (measured); $\hat{y}$, predictor output (computable); e_s, prediction error (computable); θ, vector of estimated parameters.

- Identification of disturbances and measurement noise which makes it possible to better identify the process itself,

This chapter only deals with the identification of discrete mono-variable systems LTI of the form

$$y_p(k) + \sum_{i=1}^{n} a_i y_p(k-i) = \sum_{j=0}^{m} b_j u(k-d-j) \quad k = 0, 1, 2, 3, \ldots \tag{3.1}$$

$$Y_p(z)\left[1 + a_1 z^{-1} + \cdots + a_n z^{-n}\right] = U(z) z^{-d}\left[b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}\right] \tag{3.2}$$

From where

$$G(z) = \frac{Y_p(z)}{U(z)} = z^{-d}\frac{b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}}{1 + a_1 z^{-1} + \cdots + a_n z^{-n}} = z^{-d}\frac{B(z)}{A(z)} \tag{3.3}$$

3.2 Delay operator (q^{-1})

In order to simplify the writing of the difference equations, we introduce the delay operator defined as follows:

$$\begin{aligned} q^{-i} y(k) &= y(k-i) & k \geq i \\ q^{-i} y(k) &= 0 & 0 \leq k < i \end{aligned}$$

The system equation is written as follows:

$$A\left(q^{-1}\right) y(k) = B\left(q^{-1}\right) u(k-d) = q^{-d} B\left(q^{-1}\right) u(k) \tag{3.4}$$

With

$$\begin{aligned} A\left(q^{-1}\right) &= 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \\ B\left(q^{-1}\right) &= b_0 + b_1 q^{-1} + \cdots + b_m q^{-m} \end{aligned} \tag{3.5}$$

We can thus define **the time domain transfer function** $G\left(q^{-1}\right)$, which to the discrete signal $u(k)$, corresponds to the discrete signal $y(k)$

$$G\left(q^{-1}\right)=\frac{y(k)}{u(k)}=q^{-d} \frac{B\left(q^{-1}\right)}{A\left(q^{-1}\right)} \tag{3.6}$$

We use the notation $n(k)$ to represent the discrete contribution of noise to the output $y(k)$, embodying factors such as measurement noise and disturbances. The identification process can be approached differently based on the nature of the noise:

Deterministic approach: In deterministic systems, the output is entirely determined by past and present inputs, and measurement noises can be considered negligible. Consequently, estimated parameters are also deterministic and can be derived from a reduced set of measurements.

Stochastic approach: Stochastic systems yield random output signals with unpredictable variations. When dealing with stochastic systems, estimated parameters become random variables and necessitate statistical characterization. Typically, a larger number of measurements is required compared to the deterministic approach.

3.3 Process Model

Generally, to calculate the transfer function, there are structural choices to be specified before identification:

- The number of coefficients of the polynomial $A\left(q^{-1}\right)$ (value of n),
- The number of coefficients of the polynomial $B\left(q^{-1}\right)$ (value of $m+1$),
- The delay (value of $d \geq 1$).

3.4 Performance Criteria Based on Prediction Error

The prediction error $\varepsilon(k)$ is the difference between the measured output $y(k)$ and the predicted output $\hat{y}(k)$. The predicted output at time k is calculated from all available information up to time $k-1$. This output can be presented as (Figure (3.2)):

The output is given by

$$\hat{y}(k)=F(\theta, u(k-1), u(k-2), \ldots, y(k-1), y(k-2), \ldots, \hat{y}(k-1), \hat{y}(k-2), \ldots) \tag{3.7}$$

Where F is a function to be defined. In order to express the quality of a proposed model, we consider the sum of quadratic output errors. To determine the best model, we then minimize this criterion with respect to the parameter vector $\theta=\left[a_1, \ldots, a_n, b_0, \ldots, b_m\right]^T$:

$$\min_\theta J(\theta)=\sum_{k=1}^N \varepsilon^2(k) \tag{3.8}$$

Where N represents the number of measurements available.

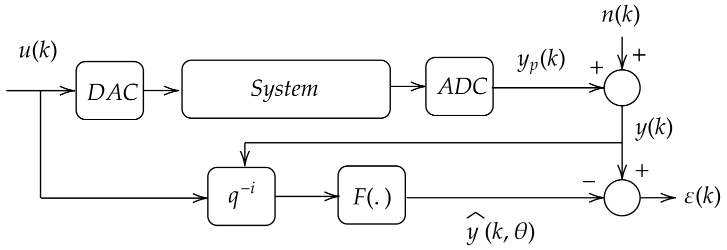

Figure 3.2: Prediction error $\varepsilon(k)$

Given the structure of the candidate model and the ultimate application objectives, a particular format for the prediction error values is necessary (e.g. equation error). This format is tailored to be minimized through the optimization methods discussed in this chapter. The fundamental error metric employed in all subsequent identification techniques is known as the prediction error. It serves as the basis for constructing the cost function, which will ultimately undergo minimization. In the following we define the prediction error and the cost function to be minimized [1].

3.5 Least squares algorithm

Consider the prediction error in the equation (3.7)

$$\min_{\theta} J(\theta) = \sum_{k=1}^{N} \varepsilon^2(k) = \sum_{k=1}^{N} (y(k) - \hat{y}(k))^2 \tag{3.9}$$

Where $y(k)$ is measured output coming from the sensor, and $\hat{y}(k)$ is the predicted output calculated from the predictor function defined above F. In this algorithm we define F as follows

$$\begin{aligned}\hat{y}(k) = F(.) = &- a_1 y(k-1) - \cdots - a_n y(k-n) + b_0 u(k-d) \\ &+ b_1 u(k-d-1) + \cdots + b_m u(k-d-m)\end{aligned} \tag{3.10}$$

It is clear that the predicted output $\hat{y}(k,\theta)$ is function of time k and the parameters vector θ. As the samples $u(k-d), \ldots, u(k-d-m)$ are given and the samples $y(k), \ldots, y(k-n)$ are measured, the prediction error $\varepsilon(k)$ is linear with respect to the parameters θ. That is for any two non zero scalars $\alpha,\ \beta$

$$\hat{y}(k, \alpha\theta_1 + \beta\theta_2) = \alpha\ \hat{y}(k,\theta_1) + \beta\ \hat{y}(k,\theta_2) \tag{3.11}$$

In this case the prediction error is known as the **equation error** and denoted by $\varepsilon(k) = \varepsilon_e(k)$

$$\begin{aligned}\varepsilon_e(k) = {} & y(k) + a_1 y(k-1) + \cdots + a_n y(k-n) - b_0 u(k-d) \\ & - b_1 u(k-d-1) - \cdots - b_m u(k-d-m) \\ = {} & y(k) - [-a_1 y(k-1) - \cdots - a_n y(k-n) + b_0 u(k-d) \\ & + b_1 u(k-d-1) + \cdots + b_m u(k-d-m)] \\ & = A\left(q^{-1}\right) y(k) - q^{-d} B\left(q^{-1}\right) u(k)\end{aligned} \tag{3.12}$$

Consider the case of minimizing the quadratic criterion based on the equation error using the least squares method. Suppose the system is strictly causal ($d \geq 1$). From the equation (3.12), we have

$$\varepsilon(k) = y(k) - \phi^T(k)\theta \tag{3.13}$$

where :

$$\begin{aligned}\phi^T(k) &= [-y(k-1), \ldots, -y(k-n), u(k-d), \ldots, u(k-d-m)] \\ \theta^T &= [a_1, \ldots, a_n, b_0, \ldots, b_m]\end{aligned} \tag{3.14}$$

where $\phi(k)$ represents the regressor, θ the vector of parameters to identify and $\varepsilon(k)$ the equation error. By accumulating N measurements, for example at discrete times $k = 1, 2, \ldots N$. The equation (3.13) gives in matrix form:

$$\begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(N) \end{bmatrix} = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} - \begin{bmatrix} \phi^T(1) \\ \vdots \\ \phi^T(N) \end{bmatrix} \theta \tag{3.15}$$

That we note

$$\mathcal{E} = Y - \Psi\theta \tag{3.16}$$

With the following dimensions:

$$\begin{array}{lcll} \mathcal{E} & : & (N \times 1) & \text{error vector} \\ Y & : & (N \times 1) & \text{vector of measurements} \\ \Psi & : & (N \times p) & \text{observation matrix} \\ \theta & : & (p \times 1) & \text{parameter vector} \end{array}$$

where $p = n + m + 1$ represents the number of parameters to identify. Note that

$$\phi^T(1) = [-y(0), \ldots, -y(1-n), u(1-d), \ldots, u(1-d-m)] \tag{3.17}$$

Elements with negative time indices will be chosen as zero. The minimization of the quadratic prediction error is written as

$$\begin{aligned}\min_{\theta} J(\theta) &= \sum_{k=1}^{N} \varepsilon^2(k) = \mathcal{E}^T\mathcal{E} \\ &= [Y-\theta]^T[Y-\Psi\theta] = Y^TY - 2Y^T\Psi\theta + \theta^T\Psi^T\Psi\theta\end{aligned} \tag{3.18}$$

The parameter vector $\hat{\theta}$ which minimizes J, cancels the gradient of J with respect to θ

$$\left.\frac{\partial J(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = -2\Psi^TY + 2\Psi^T\Psi\hat{\theta} = 0 \tag{3.19}$$

From which we get:

$$\hat{\theta} = \left(\Psi^T\Psi\right)^{-1}\Psi^TY \tag{3.20}$$

Which we also write:

$$\hat{\theta} = \Psi^+Y \tag{3.21}$$

where Ψ^+ represents the pseudo-inverse matrix to the left of Ψ. The equation (3.20) represents a necessary, but not sufficient, condition for the minimization of J. The sufficient condition requires that Ψ be of rank p and that the Hessian of J with respect to θ, evaluated at $\hat{\theta}$, is a positive definite symmetric matrix

$$\left.\frac{\partial^2 J(\theta)}{\partial\theta\partial\theta^T}\right|_{\theta=\hat{\theta}} = 2\Psi^T\Psi > 0 \tag{3.22}$$

This implies that $\hat{\theta}$ corresponds to a minimum of J. The term $\left(\Psi^T\Psi\right)$ is a square matrix of dimension p called **information matrix**. Its elements depend, among other things, on the signal $u(k)$ which excites the system. As a general rule, this matrix is regular when the input $u(k)$ is "sufficiently excited". If the matrix $\Psi^T\Psi$ is singular, the equation (3.20) has an infinite number of solutions. The estimate of the parameter vector, given by the equation (3.20), can also be written in the following equivalent form

$$\begin{aligned}\hat{\theta} &= \left[\sum_{k=1}^{N}\phi(k)\phi^T(k)\right]^{-1}\left[\sum_{k=1}^{N}\phi(k)y(k)\right] \\ &= \left[\frac{1}{N}\sum_{k=1}^{N}\phi(k)\phi^T(k)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\phi(k)y(k)\right]\end{aligned} \tag{3.23}$$

The equation (3.23) indicates that $\hat{\theta}$ can also be interpreted as the result of a correlation analysis if the number of data points N tends to infinity:

$$\hat{\theta} = R_{\phi\phi}^{-1}(0)R_{\phi y}(0) \tag{3.24}$$

3.5.1 Recursive least squares

For implementation on a computer, the equation (3.20) can be transformed into recurrent form. Considering the measurements up to time kT and with the following notations similar to those of the equation (3.15):

$$\varepsilon_k = \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(k) \end{bmatrix} Y_k = \begin{bmatrix} y(1) \\ \vdots \\ y(k) \end{bmatrix} \Psi_k = \begin{bmatrix} \phi^T(1) \\ \vdots \\ \phi^T(k) \end{bmatrix} \tag{3.25}$$

The equation (3.23) gives:

$$\hat{\theta}_k = \left[\sum_{i=1}^{k} \phi(i)\phi^T(i)\right]^{-1} \sum_{i=1}^{k} \phi(i)y(i) \tag{3.26}$$

The vector $\hat{\theta}_k$ represents the estimate of θ at time kT based on the first k equations. The goal is to calculate $\hat{\theta}_k$ from $\hat{\theta}_{k-1}$. To do this we rewrite the equation (3.26) as follows

$$\hat{\theta}_k = P_k \sum_{i=1}^{k} \phi(i)y(i) \tag{3.27}$$

where the square matrix P_k of dimension p (number of parameters) is the inverse of the information matrix.

$$P_k = \left[\sum_{i=1}^{k} \phi(i)\phi^T(i)\right]^{-1} \tag{3.28}$$

Note also that the matrix P_k^{-1} be calculated recurrently as follows:

$$\begin{aligned} P_k^{-1} &= \sum_{i=1}^{k} \phi(i)\phi^T(i) = \sum_{i=1}^{k-1} \phi(i)\phi^T(i) + \phi(k)\phi^T(k) \\ &= P_{k-1}^{-1} + \phi(k)\phi^T(k) \end{aligned} \tag{3.29}$$

The vector of parameters at time k can be written:

$$\begin{aligned} \hat{\theta}_k &= P_k \sum_{i=1}^{k} \phi(i)y(i) = P_k \left[\sum_{i=1}^{k-1} \phi(i)y(i) + \phi(k)y(k)\right] \\ &= P_k \left[P_{k-1}^{-1}\hat{\theta}_{k-1} + \phi(k)y(k)\right] \\ &= P_k \left[P_k^{-1} - \phi(k)\phi^T(k)\right] \hat{\theta}_{k-1} + P_k\phi(k)y(k) \\ &= \hat{\theta}_{k-1} + P_k\phi(k) \left[y(k) - \phi^T(k)\hat{\theta}_{k-1}\right] \end{aligned} \tag{3.30}$$

The least squares algorithm can be presented by the following two recurring equations

$$\begin{aligned} P_k^{-1} &= P_{k-1}^{-1} + \phi(k)\phi^T(k) \\ \hat{\theta}_k &= \hat{\theta}_{k-1} + P_k\phi(k) \left[y(k) - \phi^T(k)\hat{\theta}_{k-1}\right] \end{aligned} \tag{3.31}$$

At time k, we measure $y(k)$ and construct the vector $\phi(k)$. Then, we calculate P_k^{-1} and $\hat{\theta}_k$ from (3.31). To avoid the inversion of the matrix P_k at each iteration, we can use the matrix inversion lemma. Let A, C and $C^{-1} + DA^{-1}B$ be invertible matrices, then:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left[C^{-1} + DA^{-1}B\right]^{-1} DA^{-1} \tag{3.32}$$

Let $A = P_{k-1}^{-1}, B = \phi(k), C = 1, D = \phi^T(k)$ in the equation (**??**), we obtain :

$$P_k = P_{k-1} - \frac{P_{k-1}\phi(k)\phi^T(k)P_{k-1}}{1 + \phi^T(k)P_{k-1}\phi(k)} \tag{3.33}$$

This equation has the advantage of not requiring the inversion of a matrix, but only that of a scalar. There are two ways to initialize the recurrence:

- Initial values are set, generally $\hat{\theta}_0 = 0$ and $P_0 = \alpha I$ where α represents a very large scalar (e.g. $\alpha = 10000$) and I the unit matrix,

Recurring identification offers the following benefits:

- Estimation of the model in real time,
- Significant data compression, because the recurrent algorithm processes a single input/output pair at each moment instead of the entire data set,
- Lower memory requirements and computing power,

3.6 Weighted Least Squares

3.6.1 Principle

With N measurements at discrete times $k = 1, 2, \ldots, N$, the prediction error is written:

$$\mathcal{E} = Y - \Psi\theta \tag{3.34}$$

This matrix equation has N lines. All prediction errors $\varepsilon(1), \ldots, \varepsilon(N)$ do not necessarily have to have the same importance in the quadratic criterion. It is possible to weight them, for example as follows:

$$\mathcal{E}_w \equiv W[Y - \Psi\theta] \tag{3.35}$$

where W is a weighting matrix of dimension $(N \times N)$. We often choose a diagonal weighting matrix, which allows us to write:

$$\mathcal{E}_w(i) = w_{ii}\varepsilon(i) \qquad i = 1, \ldots, N$$

The following quadratic criterion is then minimized:

$$J(\theta) = \mathcal{E}_w^T\mathcal{E}_w = \mathcal{E}^T W^T W \mathcal{E} = [Y - \Psi\theta]^T W^T W[Y - \Psi\theta] \tag{3.36}$$

The vector of estimated parameters then becomes:

$$\hat{\theta} = \left(\Psi^T W^T W \Psi\right)^{-1} \Psi^T W^T W Y \tag{3.37}$$

3.6.2 Choice of weighting matrix

A selection criterion indicates that older measurements have a reduced influence in the calculation of current parameters. The influence of older measurements is artificially reduced by the introduction of the forgetting factor λ, for example:

$$W^T W = \begin{bmatrix} \lambda^{N-1} & & & 0 \\ 0 & \ddots & & \\ & & \lambda^1 & \\ 0 & & & \lambda^0 \end{bmatrix} \tag{3.38}$$

with $\lambda \leq 1$ (often $0.9 \leq \lambda \leq 0.99$)

3.6.3 Recursive weighted least squares

From the equation (3.37), we obtain:

$$\hat{\theta}_k = \left[\sum_{i=1}^{k} \phi(i)\lambda^{k-i}\phi^T(i)\right]^{-1} \sum_{i=1}^{k} \phi(i)\lambda^{k-i}y(i) = P_k \sum_{i=1}^{k} \phi(i)\lambda^{k-i}y(i) \tag{3.39}$$

The matrix P_{k+1} is calculated recurrently as follows:

$$P_k^{-1} = \lambda P_{k-1}^{-1} + \phi(k)\phi^T(k) \tag{3.40}$$

By following the same approach for the recursive least squares algorithm and the matrix inversion lemma, we obtain the following formulation with the forgetting factor:

$$\begin{aligned} P_k &= \frac{1}{\lambda}\left[P_{k-1} - \frac{P_{k-1}\phi(k)\phi^T(k)P_{k-1}}{\lambda + \phi^T(k)P_{k-1}\phi(k)}\right] \\ \hat{\theta}_k &= \hat{\theta}_{k-1} + P_k\phi(k)\left[y(k) - \phi^T(k)\hat{\theta}_{k-1}\right] \end{aligned} \tag{3.41}$$

3.7 Estimation error and variance

One of the ideal "expectations" of an estimator is that it provides us with accurate estimates. We define

Definition

An estimator $\hat{\theta}$ is said to be accurate or unbiased if and only if

$$\mu_{\hat{\theta}} = E(\hat{\theta}) = \theta_0$$

The difference $(\hat{\theta}) - \theta_0$ is said to be the **bias** of that estimator.

A more important requirement of an estimator is that it is able to deliver estimates with as less variability

Definition

The variance of an estimator (estimate) is defined as

$$\sigma^2_{\hat{\theta}} = E\left(\left(\hat{\theta} - \mu_{\hat{\theta}}\right)^2\right)$$

Observe that the definition is with reference to the average of the estimator, $\mu_{\hat{\theta}}$ and not with respect to its true value, θ_0.

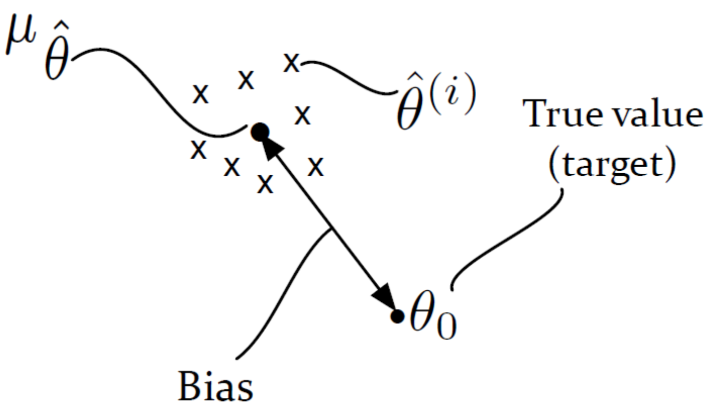

FIGURE 3.3: Illustration of the bias of the estimators [2].

Consider the linear regression problem for minimizing the error equation. It is assumed that the process can be represented by the following true model

$$A_0\left(q^{-1}\right) y(k) = q^{-d_0} B_0\left(q^{-1}\right) u(k) + e_0(k) \tag{3.42}$$

With the (real) polynomials $A_0\left(q^{-1}\right)$ and $B_0\left(q^{-1}\right)$ and the (real) noise equation $e_0(k)$. The output of this model is written:

$$\begin{aligned} y(k) =& - a_1^0 y(k-1) - \cdots - a_{n_0}^0 y\left(k - n_0\right) + b_0^0 u\left(k - d_0\right) + \cdots \\ & + b_{m_0}^0 u\left(k - d_0 - m_0\right) + e_0(k) \\ y(k) =& \phi^T(k)\theta_0 + e_0(k) \end{aligned} \tag{3.43}$$

where θ_0 is the vector of true parameters. We can express the parameters of the process model in the following regressive vector form:

$$Y = \Psi\theta_0 + E_0 \tag{3.44}$$

with $E_0^T = [e_0(1), \dots, e_0(N)]$. The prediction error in matrix form is written:

$$\mathcal{E} = Y - \Psi\theta \tag{3.45}$$

where θ is the vector of parameters to identify and $\mathcal{E}$ the vector of prediction errors to minimize. For the case of correct model structure (i.e. $n = n_0$,$m = m_0$ and $d = d_0$), combining these two last equations, we can explain the vector of prediction errors as follows:

$$\mathcal{E} = Y - \Psi\theta = (\Psi\theta_0 + E_0) - \Psi\theta = \Psi(\theta_0 - \theta) + E_0 \tag{3.46}$$

We then minimize this quadratic prediction error to obtain the parameter vector θ. We can then define the estimation error as follows:

$$\begin{aligned} \hat{\theta} &= \left(\Psi^T\Psi\right)^{-1}\Psi^T Y = \left(\Psi^T\Psi\right)^{-1}\Psi^T(\Psi\theta_0 + E_0) = \theta_0 + \left(\Psi^T\Psi\right)^{-1}\Psi^T E_0 \\ \hat{\theta} - \theta_0 &= \left[\frac{1}{N}\sum_{k=1}^{N}\phi(k)\phi^T(k)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\phi(k)e_0(k)\right] \end{aligned} \tag{3.47}$$

As E_0 is a vector of random variables, so will the vector of estimated parameters $\hat{\theta}$. If the number of data N tends towards infinity, the average value of the estimation error or bias is equal to:

$$E\left\{\hat{\theta} - \theta_0\right\} = R_{\phi\phi}^{-1}(0)R_{\phi e_0}(0) \tag{3.48}$$

Thus, the bias will be zero if the following two conditions are satisfied:

- The matrix $R_{\phi\phi}(0)$ is regular,
- $R_{\phi e_0}(0) = 0$

Condition (1) is met if the system is "sufficiently" excited. Condition (2) requires that the equation noise $e_0(k)$ of the (real) system is not correlated with the regressor $\phi(k)$. The covariance matrix of the estimated parameters can be determined from the equation (3.47)

$$\text{cov}(\hat{\theta}) = E\left\{\left(\hat{\theta} - \theta_0\right)\left(\hat{\theta} - \theta_0\right)^T\right\} = E\left\{\left(\Psi^T\Psi\right)^{-1}\Psi^T E_0 E_0^T \Psi\left(\Psi^T\Psi\right)^{-1}\right\} \tag{3.49}$$

In the particular case where the equation noise $e_0(k)$ is white we have $E\left\{E_0E_0^T\right\} = \sigma_0^2 I$, where σ_0^2 is the noise variance $e_0(k)$ and I is the unit matrix. We therefore obtain:

$$\text{cov}(\hat{\theta}) = \sigma_0^2 E\left\{\left(\Psi^T\Psi\right)^{-1}\right\} = \frac{\sigma_0^2}{N}E\left\{\left[\frac{1}{N}\sum_{k=1}^{N}\phi(k)\phi^T(k)\right]^{-1}\right\} \tag{3.50}$$

When the number of data tends towards infinity the covariance matrix is written:

$$\text{cov}(\hat{\theta}) = \frac{\sigma_0^2}{N} R_{\phi\phi}^{-1}(0) \tag{3.51}$$

The variance of the parameters depends proportionally to the variance of the noise. To have a better estimate of the parameters, a richer excitation signal must be chosen so that the information matrix becomes larger. We can always increase the number of data N to reduce the variance of the parameters.

Generally speaking, the least squares method based on the equation error is simple to implement (linear regression) but often gives biased estimates because the equation error is correlated with the regressor. It is the elimination of bias in the presence of errors (disturbances, measurement noise, model errors) which is at the origin of the development of most identification methods. These fall into two categories, depending on the approach chosen to reduce bias:

- Methods based on **the uncorrelation** of an auxiliary observation vector and the prediction error: instrumental variables method,
- Methods based on the introduction of a noise model and **minimization of the prediction error**.

3.8 Instrumental variables method

In the linear least squares parameter identification approach, the estimation error vanishes when ($R_{\phi e_0}(0) = 0$). One solution consists of creating an auxiliary observation vector ϕ_{IV} which is, by construction, uncorrelated with the noise so as to also obtain $R_{\phi_{IV} e_0}(0) \cong 0$. The instrumental variable method implements this solution. It is fundamentally based on correlation techniques and is therefore based on statistical considerations whose validity requires a large number of measurements ($N \to \infty$). The idea is to replace the linear least squares solution (3.20) with:

$$\hat{\theta}_{IV} = \left(\Psi_{\text{IV}}^{\text{T}} \Psi\right)^{-1} \Psi_{\text{IV}}^{\text{T}} Y \tag{3.52}$$

Where Ψ_{IV} is a matrix of instrumental variables (or, equivalently, ϕ_{IV} is a vector of instrumental variables). A correlation analysis indicates that the estimation error will be zero if the following two conditions are satisfied:

- The matrix $\Psi_{IV}^T \Psi$ is regular.
- $R_{\phi_{IV} e_0}(0) = 0$

Which means that the instrumental variables must be correlated with the regressor ϕ but not correlated with the noise e_0. For a finite number of data, $R_{\phi_{IV} e_0}(0)$ is not zero but it will be very small. It follows that the magnitude of the bias will also depend on the magnitude of the matrix $\Psi_{IV}^T \Psi$. The best vector of instrumental variables is the one which makes the matrix $\Psi_{IV}^T \Psi$ as large as possible. For this, $\phi_{IV}(k)$ must be a noise-free approximation of the regressor $\phi(k)$. Several possibilities exist for the choice of instrumental variables, in particular:

Shifting the noisy variables in the regressor ϕ by h

$$\phi_{IV}^T(k) = [-y(k-h-1)\ldots - y(k-h-n) \;\; u(k-d-1)\ldots u(k-d-m)] \tag{3.53}$$

Where h is chosen sufficiently large. This approach is only applicable if the noise (disturbances, measurement noise) is of high frequency compared to the bandwidth of the process and the sampling period is not too small.

Construction of instrumental variables using an auxiliary model

$$\phi_{IV}^T(k) = [-y_m(k-1)\ldots - y_m(k-n) \;\; u(k-d-1)\ldots u(k-d-m)] \tag{3.54}$$

where $y_m(k)$ is an approximation of $y_p(k)$ (the noise-free output) generated from an auxiliary model whose input is $u(k)$. As it is necessary to have a sufficiently representative auxiliary model, this approach is often initialized with the solution obtained from linear least squares.

3.9 Prediction error methods

The parametric identification problem can be studied in a more general framework with the prediction error method. This method is based on the following three steps:

1. Choose the structure of the system model. The choice of structure depends on assumptions about the order of the process model and the nature of the noise. Two different hypotheses for noise are considered:
    - Noise is not correlated with process input. This hypothesis leads to **noise model-free structures**.
    - The noise on the output is white noise filtered by a finite order filter. This hypothesis leads to **structures with noise model**.
2. Define the predictor $\hat{y}(k) = F(\theta, y(k-1), \ldots, u(k-1), \ldots)$ of the output in relation to the structure of the model.
3. Minimize the prediction error $\varepsilon(k) = y(k) - \hat{y}(k)$ with a numerical method:

$$\hat{\theta} = \arg\min_{\theta} J(\theta) = \frac{1}{N}\sum_{k=1}^{N} \varepsilon^2(k) \tag{3.55}$$

   This criterion can be minimized using the iterative Gauss-Newton algorithm which gives us a local minimum of the criterion.

3.9.1 Structures without noise model

OE structure (Output Error)

Assume that the measured output of the system can be expressed as follows:

$$y(k) = G_0\left(q^{-1}\right) u(k) + n(k) \tag{3.56}$$

where $n(k)$ is a zero-mean stationary noise independent of the input, and $G_0\left(q^{-1}\right)$ is the true model of the system:

$$G_0 = \frac{q^{-d_0} B_0\left(q^{-1}\right)}{A_0\left(q^{-1}\right)} \tag{3.57}$$

With

$$\begin{aligned} B_0\left(q^{-1}\right) &= b_0^0 + b_1^0 q^{-1} + \cdots + b_{m0}^0 q^{-m_0} \\ A_0\left(q^{-1}\right) &= 1 + a_1^0 q^{-1} + \cdots + a_{n0}^0 q^{-n_0} \\ \theta_0^T &= \left[a_1^0, \ldots a_{n_0}^0, b_0^0, \ldots, b_{m_0}^0\right] \end{aligned}$$

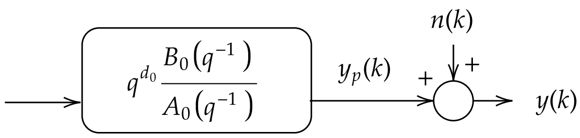

Figure 3.4: Structure *OE*

The model structure shown in equation (3.56) and Figure (3.4) is known as the output error (OE) structure. Consider the equation (3.56) and suppose that we are at time $k-1$ and that we want to predict the output of the system at time k (one-step predictor). The measured output can be written as $y(k) = y_p(k) + n(k)$. It is obvious that the output of the process model $y_p(k)$ can be easily calculated with the information available at time $k-1$ provided that the parameters of the true model $G_0\left(q^{-1}\right)$ are perfectly known. On the other hand, being a random signal, the noise $n(k)$ is not predictable before its occurrence at time k. So the best predictor for the output error structure is:

$$\hat{y}\left(k, \theta_0\right) = G_0\left(q^{-1}\right) u(k) \tag{3.58}$$

where $\hat{y}\left(k, \theta_0\right)$ is the predicted output at time k which depends on the unknown vector θ_0. We can construct a parameterized predictor where the vector θ_0 is replaced by an unknown parameter vector θ as follows:

$$\hat{y}(k, \theta) = G\left(q^{-1}, \theta\right) u(k) = \frac{q^{-d} B\left(q^{-1}\right)}{A\left(q^{-1}\right)} u(k) \tag{3.59}$$

where $G\left(q^{-1}, \theta\right)$ is the set of candidate models for the true model with:

$$\begin{aligned} B\left(q^{-1}\right) &= b_0 + b_1 q^{-1} + \cdots + b_m q^{-m} \\ A\left(q^{-1}\right) &= 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \\ \theta^T &= [a_1, \ldots a_n, b_0, \ldots, b_m] \end{aligned} \tag{3.60}$$

If $d_0 = d, m_0 = m$ and $n_0 = n$, we say that $G_0\left(q^{-1}\right)$ belongs to the set of models $G\left(q^{-1}\right)$, i.e. $G\left(q^{-1}, \theta_0\right) = G_0\left(q^{-1}\right)$. The block diagram of the predictor and the system in OE identification scheme is illustrated in the following Figure

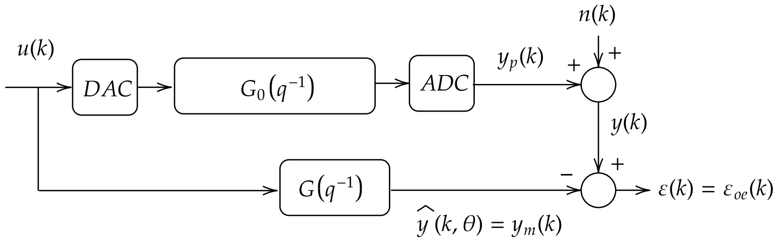

FIGURE 3.5: Structure of *OE* predictor with the process

The predicted output is indeed the model output y_m and the prediction error for this structure is the output error $\varepsilon(k,\theta) = \varepsilon_{oe}(k)$. The prediction error is written:

$$\begin{aligned} \varepsilon(k,\theta) =& \varepsilon_{oe}(k) = \; y(k) - \hat{y}(k,\theta) = G_0\left(q^{-1}\right) u(k) + n(k) - G\left(q^{-1}\right) u(k) \\ &= \left[G_0\left(q^{-1}\right) - G\left(q^{-1}\right)\right] u(k) + n(k) \end{aligned} \tag{3.61}$$

From this equation, we observe that if $G_0\left(q^{-1}\right)$ belongs to the set of models and $\theta = \theta_0$, that is. $G\left(q^{-1}\right) = G_0\left(q^{-1}\right)$, the prediction error will be equal to the noise $\varepsilon\left(k, \theta_0\right) = \varepsilon_{oe}(k) = n(k)$. Then, for $\hat{\theta} = \theta_0$, the prediction error will not be correlated with the process input. This interesting property can be used for validation of the identified model.

FIR structure

A special case of the OE structure is the case where $n_0 = 0$ or $A_0\left(q^{-1}\right) = 1$. This is called the finite impulse response (FIR) structure.

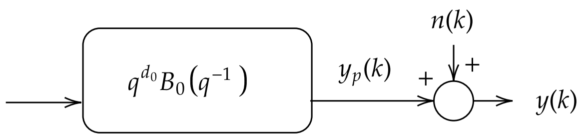

FIGURE 3.6: FIR structure For this structure we have:

$$y(k) = q^{-d} B_0 \left(q^{-1}\right) u(k) + n(k) \tag{3.62}$$

The prediction of the output of this model is written:

$$\hat{y}(k) = q^{-d} B \left(q^{-1}\right) u(k) \tag{3.63}$$

We then calculate the prediction error:

$$\varepsilon(k) = y(k) - \hat{y}(k) = y(k) - \phi_f^T(k)\theta \tag{3.64}$$

Or

$$\begin{aligned} \phi_f^T(k) &= [u(k-d), u(k-d-1), \ldots, u(k-d-m)] \\ \theta^T &= [b_0, b_1, \ldots, b_m] \end{aligned} \tag{3.65}$$

Which gives a linear regression problem, or we can use the least squares algorithm (3.20). The order of polynomial $B\left(q^{-1}\right)$ must be chosen such that $(m+d)T$ is greater than or equal to the establishment time of the impulse response of the system. The parameter estimation error is calculated as follows:

$$\hat{\theta} - \theta_0 = \left[\sum_{k=1}^{N} \phi_f(k)\phi_f^T(k)\right]^{-1} \left[\sum_{k=1}^{N} \phi_f(k) n(k)\right] \tag{3.66}$$

The parameter estimation is therefore unbiased because $n(k)$ is not correlated with the regressor $\phi_f(k)$ which only contains the input signal.

3.9.2 Structures with noise model

It can be demonstrated that if the noise spectrum is strictly positive in all frequencies between 0 and π $(\phi_{nn}(\omega) > 0, \forall\omega \in [0\pi])$, the noise $n(k)$ can always be modeled by a white noise with zero mean $e(k)$ filtered by a stable transfer function which is inversely stable:

$$n(k) = H_0\left(q^{-1}\right) e(k) = \frac{C_0\left(q^{-1}\right)}{D_0\left(q^{-1}\right)} e(k) \tag{3.67}$$

Where C_0 and D_0 are monic polynomials (first coefficient is 1) with all roots inside the unit circle. By introducing the noise model into the equation (3.67), we obtain a broader class of structures for the real system:

$$y(k) = G_0\left(q^{-1}\right) u(k) + H_0\left(q^{-1}\right) e(k) \tag{3.68}$$

As the noise at time k is filtered white noise, it also depends on the past values of e (therefore available) at time $k-1$. To find the predictable elements of noise, it is expressed as follows:

$$\begin{aligned} n(k) &= H_0\left(q^{-1}\right) e(k) = \left[H_0\left(q^{-1}\right) - 1\right] e(k) + e(k) \\ &= \left[\frac{C_0\left(q^{-1}\right) - D_0\left(q^{-1}\right)}{D_0\left(q^{-1}\right)}\right] e(k) + e(k) \end{aligned} \tag{3.69}$$

We observe that $n(k)$ contains two terms:

$$\left[\frac{C_0\left(q^{-1}\right)-D_0\left(q^{-1}\right)}{D_0\left(q^{-1}\right)}\right] e(k) \quad \text{and} \quad e(k) \tag{3.70}$$

The second term $e(k)$ is completely unpredictable at time $k-1$, while the first term is predictable if $H_0\left(q^{-1}\right)$ is known. This can be explained by expanding the difference of $C_0\left(q^{-1}\right)$ and $D_0\left(q^{-1}\right)$ which are polynomials moniques:

$$C_0\left(q^{-1}\right)-D_0\left(q^{-1}\right)=\left(c_1^0-d_1^0\right) q^{-1}+\left(c_2^0-d_2^0\right) q^{-2}+\cdots \tag{3.71}$$

We see that the first term of $n(k)$ in the equation (3.69) only depends on the values of $e(k-1), e(k-2), \ldots$ and is therefore accessible at time $k-1$. This allows us to define the best predictor of the output for the general structure (3.68) as follows:

$$\hat{y}\left(k, \theta_0\right)=G_0\left(q^{-1}\right) u(k)+\left[H_0\left(q^{-1}\right)-1\right] e(k) \tag{3.72}$$

Where θ_0 contains the parameters of the process model $G_0\left(q^{-1}\right)$ and the noise model $H_0\left(q^{-1}\right)$. With this ideal predictor, the prediction error will be equal to the white noise $e(k)$ and therefore also white:

$$\varepsilon\left(k, \theta_0\right)=y(k)-\hat{y}\left(k, \theta_0\right)=e(k) \tag{3.73}$$

Unfortunately, the parameter vector θ_0 is unknown and the values $e(k-1), e(k-2), \ldots$ are not measurable. We therefore simply define a parameterized predictor where θ_0 is replaced by θ and $e(k-1), e(k-2), \ldots$ are estimated by $\varepsilon(k-1), \varepsilon(k-2), \ldots$:

$$\hat{y}(k, \theta)=G\left(q^{-1}\right) u(k)+\left[H\left(q^{-1}\right)-1\right] \varepsilon(k, \theta) \tag{3.74}$$

The prediction error in this case is written:

$$\varepsilon(k, \theta)=y(k)-\hat{y}(k, \theta)=y(k)-G\left(q^{-1}\right) u(k)-\left[H\left(q^{-1}\right)-1\right] \varepsilon(k, \theta) \tag{3.75}$$

Which give :

$$\varepsilon(k, \theta)=H^{-1}\left(q^{-1}\right)\left[y(k)-G\left(q^{-1}\right) u(k)\right] \tag{3.76}$$

By replacing $y(k)$ from equation (3.68) into equation (3.76), we obtain:

$$\begin{aligned} \varepsilon(k, \theta) & =H^{-1}\left(q^{-1}\right)\left[G_0\left(q^{-1}\right) u(k)+H_0\left(q^{-1}\right) e(k)-G\left(q^{-1}\right) u(k)\right] \\ & =H^{-1}\left(q^{-1}\right)\left\{\left[G_0\left(q^{-1}\right)-G\left(q^{-1}\right)\right] u(k)+\left[H_0\left(q^{-1}\right)-H\left(q^{-1}\right)\right] e(k)\right\}+e(k) \end{aligned} \tag{3.77}$$

Thus, if G_0 and H_0 belong to the set of models G and H, for $\theta=\theta_0$ we will have $G=G_0$ and $H=H_0$ and thus $\varepsilon(k, \theta)=e(k)$ and therefore a blank prediction error. Intuitively, this important result can be explained as follows. The prediction error represents the disagreement between the measurement $y(k)$ and the prediction $\hat{y}(k, \theta)$ of the complete model (process and noise models). A good model will be able to explain both the deterministic component of the output $y(k)$ that results from the input $u(k)$ as well as the component that comes from the noise, which can be generated from white noise (initial hypothesis). Thus, the sequence of prediction errors (called residuals) will not be correlated with the input if the model of the process is correct. Moreover, it will be white by the equation (3.77) if the process and noise models are correct.

To simplify notation the θ argument in the model, the output predictor and the prediction error are subsequently omitted. But we must always remember that $\hat{y}(k)$ and $\varepsilon(k)$ are functions of θ.

With different assumptions about the noise model, different structures can be defined. We now present the output prediction for some special cases.

ARX structure

This structure models the noise $n(k)$ by white noise filtered using the denominator of the process model:

$$y(k) = \frac{q^{-d_0} B_0\left(q^{-1}\right)}{A_0\left(q^{-1}\right)} u(k) + \frac{1}{A_0\left(q^{-1}\right)} e(k) \tag{3.78}$$

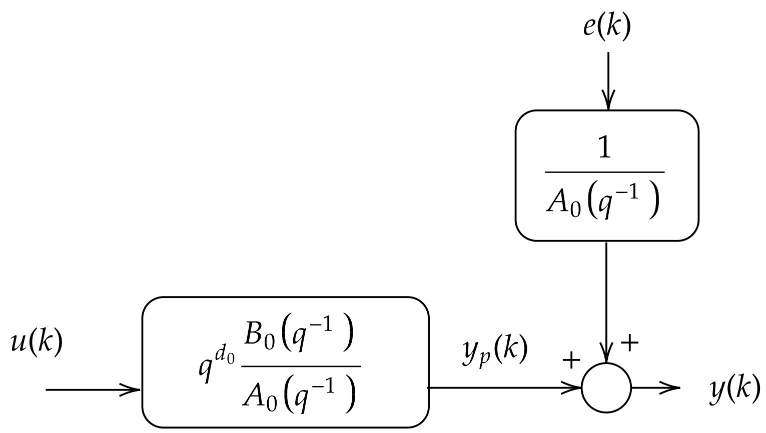

FIGURE 3.7: Structure *ARX*

This structure is known as ARX, where AR refers to the Auto-Regressive term $A_0\left(q^{-1}\right) y(k)$ (when we multiply both sides of the equation (3.78) by $A_0\left(q^{-1}\right)$) and X to external input $u(k)$.

The output predictor for the ARX model is obtained by replacing $H\left(q^{-1}\right)$ with $\frac{1}{A(q^{-1})}$ in the equation (3.74):

$$\hat{y}(k) = \frac{q^{-d} B\left(q^{-1}\right)}{A\left(q^{-1}\right)} u(k)) + \left[\frac{1}{A\left(q^{-1}\right)} - 1\right] \varepsilon(k) \tag{3.79}$$

The prediction error is also obtained by replacing $H\left(q^{-1}\right)$ by $\frac{1}{A(q^{-1})}$ in the equation (3.76)

$$\begin{aligned}
\varepsilon(k) =& A\left(q^{-1}\right)\left[y(k) - \frac{q^{-d}B\left(q^{-1}\right)}{A\left(q^{-1}\right)}u(k)\right] \\
=& A\left(q^{-1}\right)y(k) - B\left(q^{-1}\right)u(k-d) \\
=& y(k) - a_1 y(k-1) - \cdots - a_n y(k-n) + b_0 u(k-d) + \cdots + b_m u(k-d-m) \\
=& y(k) - \phi^T(k)\theta
\end{aligned} \tag{3.80}$$

The advantage is that the predictor based on this structure is linear with respect to the model parameter vectors to be identified. We observe that the prediction error equation for the ARX model is the same as the error equation (3.12). This allows us to use the least squares algorithm for the minimization of the quadratic criterion based on the prediction error.

ARMAX structure

For this structure, the measured output is written in Figure (3.8):

$$y(k) = \frac{q^{-d_0}B_0\left(q^{-1}\right)}{A_0\left(q^{-1}\right)}u(k) + \frac{C_0\left(q^{-1}\right)}{A_0\left(q^{-1}\right)}e(k) \tag{3.81}$$

The ARMAX (AutoRegressive–Moving-Average with eXternal input) model. As the two transfer functions in figure (3.8) have the same denominator, this model is useful when noise acts upstream of the process.

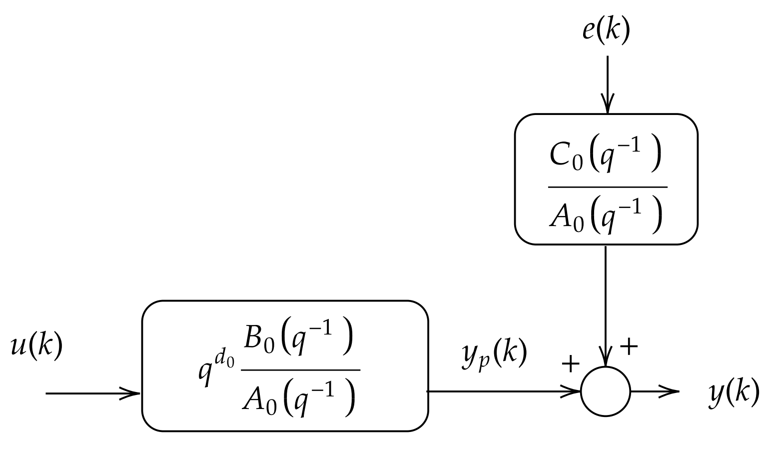

FIGURE 3.8: ARMAX structure

By replacing $H\left(q^{-1}\right)$ by $\frac{C(q^{-1})}{A(q^{-1})}$ in the equation (3.74) we obtain the prediction of the output for the ARMAX model:

$$\hat{y}(k) = \frac{q^{-d}B\left(q^{-1}\right)}{A\left(q^{-1}\right)}u(k) + \left[\frac{C\left(q^{-1}\right)}{A\left(q^{-1}\right)} - 1\right]\varepsilon(k) \tag{3.82}$$

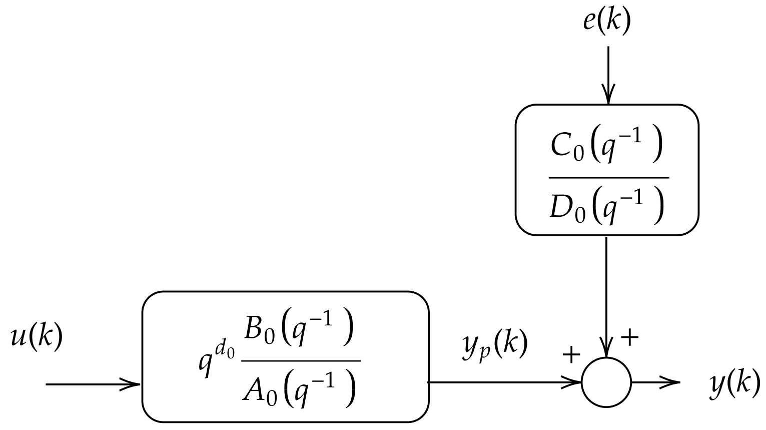

FIGURE 3.9: Box-Jenkins structure

The prediction error for the ARMAX model is therefore written:

$$\begin{aligned}\varepsilon(k) &= \frac{A\left(q^{-1}\right)}{C\left(q^{-1}\right)}\left[y(k) - \frac{q^{-d}B\left(q^{-1}\right)}{A\left(q^{-1}\right)}u(k)\right] \\ &= \frac{1}{C\left(q^{-1}\right)}\left[A\left(q^{-1}\right)y(k) - B\left(q^{-1}\right)u(k-d)\right]\end{aligned} \tag{3.83}$$

The prediction error is nonlinear with respect to the process and noise model parameters.

Box-Jenkins structure

The structure proposed by Box and Jenkins is more general because the process and noise models do not have a common factor. The measured output is represented by:

$$y(k) = \frac{q^{-d_0}B_0\left(q^{-1}\right)}{A_0\left(q^{-1}\right)}u(k) + \frac{C_0\left(q^{-1}\right)}{D_0\left(q^{-1}\right)}e(k) \tag{3.84}$$

This structure also leads to a non-linear predictor with respect to the parameter vector

$$\hat{y}(k) = \frac{q^{-d}B\left(q^{-1}\right)}{A\left(q^{-1}\right)}u(k)) + \left[\frac{C\left(q^{-1}\right)}{D\left(q^{-1}\right)} - 1\right]\varepsilon(k) \tag{3.85}$$

The prediction error for the model BJ is written from the equation (3.76) by replacing $H\left(q^{-1}\right)$ by the function $\frac{C\left(q^{-1}\right)}{D\left(q^{-1}\right)}$

$$\varepsilon(k) = \frac{D\left(q^{-1}\right)}{C\left(q^{-1}\right)}\left[y(k) - \frac{q^{-d}B\left(q^{-1}\right)}{A\left(q^{-1}\right)}u(k)\right] \tag{3.86}$$

3.9.3 Minimization of prediction error

We have seen that the parameters of a model are identified by minimizing a quadratic criterion based on the prediction error defined as

$$J(\theta) = \frac{1}{N}\sum_{k=1}^{N}\varepsilon^2(k) = \frac{1}{N}\sum_{k=1}^{N}[y(k)-\hat{y}(k)]^2 \tag{3.87}$$

For FIR and ARX structures, the prediction error is linear with respect to the vector of parameters to be identified and therefore the minimum of the criterion (3.87) can be calculated with the least squares method. The other structures give a non-linear optimization problem for which we use another approach is to solve this problem iteratively by solving at each iteration a linear regression problem hence the name pseudo-linear regression.

Pseudo-linear regression

Consider the output error predictor given in equation (3.59). By multiplying both sides of this equation by $A\left(q^{-1}\right)$ we obtain:

$$A\left(q^{-1}\right)\hat{y}(k) = B\left(q^{-1}\right)u(k-d) \tag{3.88}$$

This equation can be represented in vector form as follows:

$$\begin{aligned}\hat{y}(k) &= -a_1\hat{y}(k-1),\cdots,-a_n\hat{y}(k-n)+b_0u(k-d)+\cdots+b_mu(k-d-m)\\ &= \phi_e^T(k)\theta\end{aligned} \tag{3.89}$$

Or

$$\begin{aligned}\phi_e^T(k) &= [-\hat{y}(k-1),\ldots,-\hat{y}(k-n),u(k-d),\ldots,u(k-d-m)]\\ \theta^T &= [a_1,\ldots,a_n,b_0,\ldots,b_m]\end{aligned} \tag{3.90}$$

The prediction error $\varepsilon(k) = y(k)-\phi_e^T(k)\theta$ has the form of a linear regression but in reality it is a regression problem non-linear because the regressor $\phi_e(k)$ contains the estimated output $\hat{y}$ which is a function of θ. This problem can be solved with an iterative method with the following algorithm:

$$\hat{\theta}_i = \left[\sum_{k=1}^{N}\phi_e\left(k,\hat{\theta}_{i-1}\right)\phi_e^T\left(k,\hat{\theta}_{i-1}\right)\right]^{-1}\left[\sum_{k=1}^{N}\phi_e\left(k,\hat{\theta}_{i-1}\right)y(k)\right] \tag{3.91}$$

This approach known as the substitution algorithm is the simplest algorithm for solving non-linear optimization problems. An **initialization** of the algorithm with the **least squares** method is **essential**.

This algorithm can also be applied to structures with a noise model. As an example, consider the ARMAX structure with the prediction error presented in equation (3.83). By multiplying both sides of this equation by $C\left(q^{-1}\right)$, the prediction error can be reformulated as a pseudo-linear regression as follows:

$$\begin{aligned}\varepsilon(k) =& y(k) - [-a_1 y(k-1) - \cdots - a_n y(k-n) + b_0 u(k-d) + \cdots + b_m u(k-d-m) \\ &+ c_1 \varepsilon(k-1) + \cdots + c_p \varepsilon(k-p)] \\ =& y(k) - \phi_x^T(k)\theta\end{aligned} \tag{3.92}$$

where p is the order of the polynomial $C\left(q^{-1}\right)$ and

$$\begin{aligned}\phi_x^T(k) =& [-y(k-1), \ldots, -y(k-n), u(k-d), \ldots, u(k-d-m) \\ & \varepsilon(k-1), \ldots, \varepsilon(k-p)] \\ \theta^T =& [a_1, \ldots, a_n, b_0, \ldots, b_m, c_1, \ldots, c_p]\end{aligned} \tag{3.93}$$

An estimate of θ is obtained by replacing $\phi_e\left(k, \hat{\theta}_i\right)$ by $\phi_x\left(k, \hat{\theta}_i\right)$ in the equation (3.91).

Gauss-Newton algorithm

The pseudo-linear regression method is very simple to implement but the convergence of the algorithm cannot be demonstrated. In practice, we prefer to use the Gauss-Newton algorithm. In this algorithm, the parameter vector at iteration i is estimated with the following relationship:

$$\hat{\theta}_i = \hat{\theta}_{i-1} - \left[J''\left(\hat{\theta}_{i-1}\right)\right]^{-1} J'\left(\hat{\theta}_{i-1}\right) \tag{3.94}$$

where $J'\left(\hat{\theta}_{i-1}\right)$ is the gradient and $J''\left(\hat{\theta}_{i-1}\right)$ the Hessian of the criterion (3.87) evaluated at $\hat{\theta}_{i-1}$. The algorithm is initialized by the parameters identified with the least squares (ARX) method for the process model. The noise model parameters are often initialized to zero. The gradient and the Hessian are easily calculated for each structure. The gradient of $J(\theta)$ with respect to θ is written:

$$J'(\theta) = \frac{\partial J}{\partial \theta} = \frac{-2}{N}\sum_{k=1}^{N}\frac{\partial \hat{y}}{\partial \theta}\varepsilon(k) = \frac{-2}{N}\sum_{k=1}^{N}\rho(k)\varepsilon(k) \tag{3.95}$$

where $\rho(k)$ is the gradient of the predictor:

$$\rho(k) \equiv \frac{\partial \hat{y}}{\partial \theta} \tag{3.96}$$

The Hessian of the criterion is calculated as follows:

$$J''(\theta) = \frac{\partial^2 J}{\partial \theta \partial \theta^T} = \frac{2}{N}\sum_{k=1}^{N}\left[\rho(k)\rho^T(k) - \frac{\partial \rho}{\partial \theta}\varepsilon(k)\right] \approx \frac{2}{N}\sum_{k=1}^{N}\rho(k)\rho^T(k) \tag{3.97}$$

The second Hessian term is ignored because it is very small compared to the first term in the vicinity of the solution ($\varepsilon(k)$ approaches the white noise $e(k)$ whose sum over N data converges to zero as N tends to infinity). To implement the Gauss-Newton algorithm we need to calculate the gradient of the output prediction $\rho(k)$.

As an example, we calculate $\rho(k)$ for the output error structure with the following predictor:

$$\hat{y}(k) = \frac{q^{-d}B\left(q^{-1}\right)}{A\left(q^{-1}\right)}u(k)) = \frac{b_0 + b_1 q^{-1} + \cdots + b_m q^{-m}}{1 + a_1 q^{-1} + \cdots + a_n q^{-n}}u(k-d) \tag{3.98}$$

We calculate the derivative of $\hat{y}(k)$ for the numerator and denominator parameters separately:

$$\begin{aligned} \frac{\partial \hat{y}}{\partial b_j} &= \frac{q^{-j}}{A\left(q^{-1}\right)}u(k-d) = \frac{1}{A\left(q^{-1}\right)}u(k-d-j) \quad j = 0,\ldots,m \\ \frac{\partial \hat{y}}{\partial a_j} &= \frac{-q^{-j}B\left(q^{-1}\right)}{A^2\left(q^{-1}\right)}u(k-d) = \frac{-1}{A\left(q^{-1}\right)}\hat{y}(k-j) \quad j = 1,\ldots,n \end{aligned} \tag{3.99}$$

We therefore obtain:

$$\begin{aligned} \rho^T(k) &= \frac{1}{A\left(q^{-1}\right)}[-\hat{y}(k-1),\ldots,-\hat{y}(k-n),u(k-d),\ldots,u(k-d-m)] \\ &= \frac{1}{A\left(q^{-1}\right)}\phi_e^T(k) \end{aligned} \tag{3.100}$$

The Gauss-Newton algorithm for identifying parameters of the output error structure can be summarized as follows:

1. Choose the order of the system to identify (d, n, m) and the stopping criterion $\epsilon > 0$.
2. Initialize the parameter vector with the least squares method $(i = 0)$:
$$\hat{\theta}_0 = \left(\Psi^T\Psi\right)^{-1}\Psi Y \tag{3.101}$$
3. Calculate $\hat{y}\left(k, \hat{\theta}_{i-1}\right) = G\left(q^{-1}, \hat{\theta}_{i-1}\right)u(k)$ for $k = 1,\ldots,N$
4. Calculate prediction error $\varepsilon\left(k, \hat{\theta}_{i-1}\right) = y(k) - \hat{y}\left(k, \hat{\theta}_{i-1}\right)$ for $k = 1,\ldots,N$.
5. Calculate the vector $\rho^T\left(k, \hat{\theta}_{i-1}\right) = \frac{1}{A\left(q^{-1}, \hat{\theta}_{i-1}\right)}\phi_e^T\left(k, \hat{\theta}_{i-1}\right)$ for $k = 1,\ldots,N$ with
$$\phi_e^T\left(k, \hat{\theta}_{i-1}\right) = \left[-\hat{y}\left(k-1, \hat{\theta}_{i-1}\right),\ldots,-\hat{y}\left(k-n, \hat{\theta}_{i-1}\right), u(k-d),\ldots u(k-m)\right] \tag{3.102}$$
6. Calculate the gradient of the criterion $J'\left(\hat{\theta}_{i-1}\right) = \frac{-2}{N}\sum_{k=1}^N \rho\left(k, \hat{\theta}_{i-1}\right)\varepsilon\left(k, \hat{\theta}_{i-1}\right)$
7. Calculate the Hessian of the criterion $J''\left(\hat{\theta}_{i-1}\right) = \frac{2}{N}\sum_{k=1}^N \rho\left(k, \hat{\theta}_{i-1}\right)\rho^T\left(k, \hat{\theta}_{i-1}\right)$
8. Calculate $\hat{\theta}_i = \hat{\theta}_{i-1} - \left[J''\left(\hat{\theta}_{i-1}\right)\right]^{-1}J'\left(\hat{\theta}_{i-1}\right)$
9. If $\left(\hat{\theta}_i - \hat{\theta}_{i-1}\right)^T\left(\hat{\theta}_i - \hat{\theta}_{i-1}\right) < \epsilon$, stop the algorithm with $\hat{\theta} = \hat{\theta}_i$ otherwise $i = i + 1$ and return to 3.

Note 3 *Recall that the measured output of a system is a random variable because of the influence of noise on the output. Likewise, the prediction error* $\varepsilon(k)$ *and the vector of estimated parameters* $\hat{\theta}$ *are* *also random variables in the sense that with the acquisition of new data we have a realization of the noise and therefore also a realization of the vector $\hat{\theta}$. The covariance of the estimated parameter vector will depend on the variance of the noise on the output σ_0^2. The parameter covariance is determined using the following relationship (when N is large enough):*

$$\operatorname{cov}(\hat{\theta}) = E\left\{\left(\hat{\theta}-\theta_0\right)\left(\hat{\theta}-\theta_0\right)^T\right\} = \frac{\sigma_0^2}{N}\left[\frac{1}{N}\sum_{k=1}^{N}\rho(k)\rho^T(k)\right]^{-1} \tag{3.103}$$

3.10 Closed Loop Identification

The identification algorithms presented above consider the system in an open-loop configuration, meaning without feedback from the output to the input. However, for reasons of stability or performance, it is often not possible to open the loop of the controlled system. Therefore, it is useful to study the performance of parametric identification algorithms in a closed-loop context. There are several methods to be adopted when identifying transfer function in the closed loop, in the next paragraph we consider one of the simplest solutions

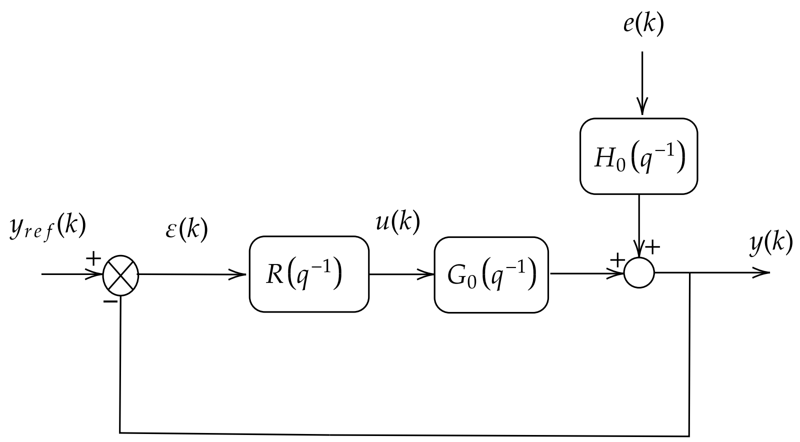

FIGURE 3.10: Closed-loop identification diagram with external excitation $y_c(k)$.

The closed-loop transfer function relating $y_c(k)$ to $y(k)$ is identified as illustrated in Figure . With the assumption that the regulator is known, the process transfer function is deduced. Therefore, we have:

$$G_{cl}\left(q^{-1}\right) = \frac{y(k)}{y_c(k)} = \frac{R\left(q^{-1}\right)G_0\left(q^{-1}\right)}{1+R\left(q^{-1}\right)G_0\left(q^{-1}\right)} \tag{3.104}$$

and

$$G_0\left(q^{-1}\right) = \frac{G_{BF}\left(q^{-1}\right)}{R\left(q^{-1}\right)\left[1-G_{BF}\left(q^{-1}\right)\right]} \tag{3.105}$$

This approach thus reduces the closed-loop identification problem of $G_0\left(q^{-1}\right)$ to the open-loop identification of $G_{BF}\left(q^{-1}\right)$. However, the model $G_{cl}\left(q^{-1}\right)$ is generally of higher order than $G_0\left(q^{-1}\right)$, which hinders parameter convergence. Furthermore, the calculation of $G_0\left(q^{-1}\right)$ from $G_{cl}\left(q^{-1}\right)$ can pose numerical difficulties, particularly related to pole/zero cancellations.

Exercise 3.10.1 ***Statement:*** *Consider the dynamic system given by the discrete transfer function:*

$$G(z) = \frac{z^3 + 2z^2 + z - 1}{z^4 + 1.6z^2 - 0.8z} \tag{3.106}$$

1. *Is this system physically realizable?*

2. *Represent this system in the form of a difference equation.*

3. *Express the transfer function in terms of negative powers of z, indicating any possible delay.*

Solution:

1. ***Physical realizability:***

   The system is physically realizable because the degree of the denominator polynomial (4) is greater than or equal to the degree of the numerator polynomial (3). This ensures that the system is causal.

2. ***Difference equation:***
   From $G(z) = \frac{Y(z)}{U(z)} = \frac{z^3+2z^2+z-1}{z^4+1.6z^2-0.8z}$, cross-multiplying gives:

   $$(z^4 + 1.6z^2 - 0.8z)Y(z) = (z^3 + 2z^2 + z - 1)U(z)$$

   Taking the inverse z-transform:

   $$y(k+4) + 1.6y(k+2) - 0.8y(k+1) = u(k+3) + 2u(k+2) + u(k+1) - u(k)$$

   Shifting back by 4 samples:

   $$y(k) + 1.6y(k-2) - 0.8y(k-3) = u(k-1) + 2u(k-2) + u(k-3) - u(k-4)$$

3. ***Transfer function with negative powers:***

   Factoring z from the denominator:

   $$G(z) = \frac{z^3 + 2z^2 + z - 1}{z(z^3 + 1.6z - 0.8)} = z^{-1}\frac{1 + 2z^{-1} + z^{-2} - z^{-3}}{1 + 1.6z^{-2} - 0.8z^{-3}}$$

   - *Pure delay: $d = 1$ sampling period*
   - *Order of denominator: $n = 3$*
   - *Order of numerator: $m = 3$*

Exercise 3.10.2 ***Statement:*** *The step response of an LTI system gave the following values for the first 3 sampling points ($T = 0.2$): $y_1 = 0.18$, $y_2 = 0.33$, and $y_3 = 0.45$. The system can be modeled by a first-order transfer function:*

$$G(z) = \frac{b_1 z^{-1}}{1 + a_1 z^{-1}} \tag{3.107}$$

Consider this numerical data for a step response:

$$u(0) = 1, \quad y(0) = 0 \quad \textit{(initial condition)}$$
$$u(1) = 1, \quad y(1) = 0.18$$
$$u(2) = 1, \quad y(2) = 0.33$$
$$u(3) = 1, \quad y(3) = 0.45$$

1. *Explain the quadratic function related to the output error minimization.*
2. *Repeat the previous development for the equation error.*
3. *Identify the dynamic system, which means determine the numerical values of* a_1 *and* b_1.
4. *Identify the dynamic system using only the first two measurements* y_1 *and* y_2.

Solution:

- ***Output error model:***

$$y_m(k) = -a_1 y_m(k-1) + b_1 u(k-1) \quad \textit{for } k = 1, 2, 3, \ldots$$

where $y_m(k)$ *is the model output.*

- ***Equation error model:***

$$\hat{y}(k) = -a_1 y(k-1) + b_1 u(k-1) \quad \textit{for } k = 1, 2, 3, \ldots$$

where $\hat{y}(k)$ *is the predicted output based on actual measurements.*

Substituting the numerical data:

$$\textit{Output error model:} \quad \begin{cases} y_m(1) = b_1 \\ y_m(2) = b_1(1 - a_1) \\ y_m(3) = b_1(1 - a_1 + a_1^2) \end{cases}$$

$$\textit{Equation error model:} \quad \begin{cases} \hat{y}(1) = b_1 \\ \hat{y}(2) = -0.18a_1 + b_1 \\ \hat{y}(3) = -0.33a_1 + b_1 \end{cases}$$

1. ***Output error minimization:***

The quadratic criterion minimizes:

$$J_{oe}(\theta) = \sum_{k=1}^{3} \varepsilon_{oe}^2(k) = \sum_{k=1}^{3} [y(k) - y_m(k, \theta)]^2$$

where $\theta = [a_1, b_1]^T$.

Substituting values:

$$J_{oe}(a_1, b_1) = [0.18 - b_1]^2 + [0.33 - b_1(1 - a_1)]^2 + [0.45 - b_1(1 - a_1 + a_1^2)]^2$$

The error $\varepsilon_{oe}(k)$ is nonlinear with respect to a_1 and b_1.

2. ***Equation error minimization:***

   The quadratic criterion minimizes:

$$J_e(\theta) = \sum_{k=1}^{3} \varepsilon_e^2(k) = \sum_{k=1}^{3} [y(k) - \hat{y}(k, \theta)]^2$$

   Substituting values:

$$J_e(a_1, b_1) = [0.18 - b_1]^2 + [0.33 + 0.18a_1 - b_1]^2 + [0.45 + 0.33a_1 - b_1]^2$$

   The error $\varepsilon_e(k)$ is linear with respect to a_1 and b_1.

3. ***System identification:***

   - ***Output error minimization:***

     Solving the system:

$$\frac{\partial J_{oe}}{\partial a_1} = 0$$
$$\frac{\partial J_{oe}}{\partial b_1} = 0$$

     This yields (approximately):

$$a_1 = -0.8191, \quad b_1 = 0.1809$$

     Transfer function:

$$\hat{G}_{oe}(z) = \frac{0.1809z^{-1}}{1 - 0.8191z^{-1}}$$

   - ***Equation error minimization:***

     Solving the linear system:

$$3b_1 - (0.18 + 0.33)a_1 = 0.18 + 0.33 + 0.45$$
$$(0.18 + 0.33)b_1 - (0.18^2 + 0.33^2)a_1 = 0.18^2 + 0.33^2 + 0.45 \times 0.33$$

     This yields (approximately):

$$a_1 = -0.819, \quad b_1 = 0.181$$

Transfer function:

$$\hat{G}_e(z) = \frac{0.181z^{-1}}{1 - 0.819z^{-1}}$$

4. ***Using only first two measurements:***

   With $y_1 = 0.18$ *and* $y_2 = 0.33$*:*

$$b_1 = 0.18$$
$$0.33 = b_1(1 - a_1) = 0.18(1 - a_1)$$

   Solving:

$$a_1 = 1 - \frac{0.33}{0.18} = -0.8333, \quad b_1 = 0.18$$

   Transfer function:

$$\hat{G}(z) = \frac{0.18z^{-1}}{1 - 0.8333z^{-1}}$$

Exercise 3.10.3 ***Statement:*** *The following measurements were made on an initially at-rest dynamic process, with the output* $y(k)$ *being disturbed by measurement noise* $n(k)$*:*

k	0	1	2	3	4	5	6	7	8	9	10
$u(k)$	0	1	-1	1	1	1	-1	-1	0	0	0
$y(k)$	0	1.1	-0.2	0.1	0.9	1	0.1	-1.1	-0.8	-0.1	0

Table 3.1: System Measurements

1. *Calculate* b_0 *and* b_1 *of the model:*

$$y(k) = b_0 u(k) + b_1 u(k-1)$$

   using the least squares method.

2. *Evaluate the noise sequence* $n(k)$*, its mean value, and its standard deviation.*

3. *Repeat part a) using the output error and equation error successively.*

Solution:

1. ***Least squares estimation:***

   The prediction error is:

$$\varepsilon(k) = y(k) - [b_0 u(k) + b_1 u(k-1)] = y(k) - \phi^T(k)\theta$$

   where $\phi(k) = [u(k), u(k-1)]^T$ *and* $\theta = [b_0, b_1]^T$*.*

For $k = 1, 2, \ldots, 10$*, construct:*

$$\Psi = \begin{bmatrix} u(1) & u(0) \\ u(2) & u(1) \\ u(3) & u(2) \\ u(4) & u(3) \\ u(5) & u(4) \\ u(6) & u(5) \\ u(7) & u(6) \\ u(8) & u(7) \\ u(9) & u(8) \\ u(10) & u(9) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \\ -1 & 1 \\ -1 & -1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$Y = \begin{bmatrix} y(1) & y(2) & y(3) & y(4) & y(5) & y(6) & y(7) & y(8) & y(9) & y(10) \end{bmatrix}^T = \begin{bmatrix} 1.1 & -0.2 & 0.1 & 0.9 & 1 & 0.1 & -$$

The least squares solution is:

$$\hat{\theta} = (\Psi^T \Psi)^{-1} \Psi^T Y = \begin{bmatrix} 0.6143 \\ 0.5143 \end{bmatrix}$$

So $b_0 = 0.6143$ *and* $b_1 = 0.5143$.

2. ***Noise analysis:***

   The residuals (estimated noise) are:

$$\hat{n}(k) = y(k) - [0.6143u(k) + 0.5143u(k-1)]$$

   Calculating for $k = 1$ *to* 10*:*

$$\hat{n} = [0.4857, -0.1, 0, -0.2286, -0.1286, 0.2, 0.0286, -0.2587, -0.1, 0]^T$$

   - *Mean:* $\bar{n} = -0.0129$
   - *Standard deviation:* $\sigma_n = 0.2225$

3. ***Output error vs. equation error:***

   For the FIR model $y(k) = b_0 u(k) + b_1 u(k-1)$*:*

   - *The output error and equation error formulations are identical*
   - *Both minimize the same quadratic criterion*
   - *Both yield the same least squares solution as in part (a)*

Exercise 3.10.4 ***Statement:*** *Given the prediction error defined by the following difference equation:*

$$\varepsilon(k) = y(k) - [-ay(k-1) + bu(k-1)]$$

Show that if the input corresponds to a proportional feedback of the output, identifying the parameters a and b using the least squares method is not possible.

Solution:

- *The prediction error can be written as:*

$$\varepsilon(k) = y(k) - [-y(k-1), u(k-1)] \begin{bmatrix} a \\ b \end{bmatrix}$$

- *With proportional feedback:* $u(k) = -gy(k)$
- *The regression matrix becomes:*

$$\Psi = \begin{bmatrix} -y(0) & u(0) \\ -y(1) & u(1) \\ -y(2) & u(2) \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} -y(0) & -gy(0) \\ -y(1) & -gy(1) \\ -y(2) & -gy(2) \\ \vdots & \vdots \end{bmatrix}$$

- *The second column is g times the first column:*

$$\textit{Column 2} = g \times \textit{Column 1}$$

- *This makes Ψ rank-deficient (columns are linearly dependent)*
- *Therefore, $\Psi^T\Psi$ is singular and not invertible*
- *The least squares solution $\hat{\theta} = (\Psi^T\Psi)^{-1}\Psi^T Y$ does not exist*
- *Only the combined parameter $(a + bg)$ can be identified from:*

$$\varepsilon(k) = y(k) + (a + bg)y(k-1)$$

Exercise 3.10.5 1. *Formulate the quadratic criteria based on the output error and the equation error to identify the dynamic system:*

$$G(z) = \frac{bz^{-3}}{1 + az^{-1}}$$

2. *Show that the criterion based on the output error generates a nonlinear regression problem, unlike the criterion based on the equation error.*

Solution:

1. ***Quadratic criteria formulation:***

   - ***Output error criterion:***
     The model output satisfies:

$$y_m(k) = -ay_m(k-1) + bu(k-3)$$

The output error is:

$$\varepsilon_{oe}(k) = y(k) - y_m(k)$$

The quadratic criterion to minimize:

$$J_{oe}(a,b) = \sum_{k=1}^{N} [y(k) + ay_m(k-1) - bu(k-3)]^2$$

with initial conditions: $y_m(0) = 0$, $u(0) = u(-1) = u(-2) = 0$.

- ***Equation error criterion:***

  The predicted output is:

  $$\hat{y}(k) = -ay(k-1) + bu(k-3)$$

  The equation error is:

  $$\varepsilon_e(k) = y(k) - \hat{y}(k) = y(k) + ay(k-1) - bu(k-3)$$

  The quadratic criterion to minimize:

  $$J_e(a,b) = \sum_{k=1}^{N} [y(k) + ay(k-1) - bu(k-3)]^2$$

  with initial conditions: $y(0) = 0$, $u(0) = u(-1) = u(-2) = 0$.

2. ***Linearity analysis:***

- ***Output error (nonlinear):***

  Expanding the first few terms:

  $$\begin{aligned} k = 1: &\quad \varepsilon_{oe}(1) = y(1) - bu(-2) \\ k = 2: &\quad \varepsilon_{oe}(2) = y(2) + a[y(1) - bu(-2)] - bu(-1) \\ k = 3: &\quad \varepsilon_{oe}(3) = y(3) + a\{a[y(1) - bu(-2)] - bu(-1)\} - bu(0) \end{aligned}$$

  The terms contain products ab, a^2b, *etc., making the error nonlinear in parameters* a *and* b.

- ***Equation error (linear):***

  Expanding the first few terms:

  $$\begin{aligned} k = 1: &\quad \varepsilon_e(1) = y(1) + ay(0) - bu(-2) \\ k = 2: &\quad \varepsilon_e(2) = y(2) + ay(1) - bu(-1) \\ k = 3: &\quad \varepsilon_e(3) = y(3) + ay(2) - bu(0) \end{aligned}$$

  All terms are linear in a *and* b, *making this a linear regression problem.*

Chapter 4

Practical Aspects and Validation of the Model

4.1 Introduction

In the field of system identification, the practical aspects related to the sampling period and model validation are crucial considerations when developing models for real-world systems. System identification involves constructing mathematical models that describe the behavior of a system based on observed input and output data. Practical constraints in identification, such as sensor limitations, data storage capacity, and real-time processing requirements, often influence the choice of the sampling period. Providing a balance between accuracy and practical constraints is crucial. In this chapter we present some of the practical aspects in identification of dynamical systems. The success of an identification depends on the following points

- The choice of the sampling period and the excitation signal $u(k)$ (amplitude and frequency content),
- The choice of process model and the noise model (number of poles and zeros, pure delay)
- The choice of the identification algorithm

4.2 Signal Conditioning

It is surprising to observe, through practical experience, the diversity of issues that can be encountered in real industrial trials :

- **Instrumentation faults:** Blocked sensors, transducer drift, and more.
- **Acquisition faults:** Random noise in the processing chain (faulty isolation, etc.).
- **Storage faults:** Partial data file corruption.
- **Actuator faults:** Random sticking of a valve, saturation due to improper calibration of the test, and so on.
- **Process incidents:** Low-frequency, small-amplitude disturbances with significant impact; unknown changes in the environment, etc.

Pre-processing of data may also be motivated by the assumptions, limitations and requirements of model development. For instance, data may contain drifts, trends and other non-stationarities, whereas most identification methods assume stationarity of data, a condition requiring the statistical properties of data to remain invariant with time.

4.2.1 Practical Guidelines for Sampling

The sampling period T must be chosen carefully before the start of the experiment, because it cannot be reduced once the samples are stored. On the other hand, the sampling period can easily be doubled, tripled, etc. by simply dropping part of the data (decimation).

An upper bound for T is given by the Shannon sampling condition ($\omega_N > \omega_{\max}$, i.e. $T < pi/\omega_{\max}$, where ω_N represents the Nyquist frequency and $\omega_{\max}$ the largest frequency contained in the input and output signals . In practice, it is recommended to use a filter in order to reduce the frequency content at high frequencies. The value of T mainly depends on:

- The dominant time constant of the final application,
- The desired precision (at low and high frequencies) of the identified model,
- Numerical difficulties that can result if T is too small.

These points are explained below.

1. We often choose the sampling period T in the following interval:

$$\frac{\tau_{cl}}{6} < T < \frac{\tau_{cl}}{2} \tag{4.1}$$

where τ_{cl} represents the dominant time constant of the looped system (or $\omega_{cl} = 1/\tau_{cl}$ its bandwidth). It should be mentioned that the closed loop bandwidth is chosen normally equal to or slightly higher than the open loop system bandwidth ω_{ol}. An estimate of the dominant time constant of the open-loop system $\tau_{ol} = 1/\omega_{ol}$ can be obtained from the step response of the system (settling time $\approx 4\tau_{ol}$).

2. If the purpose of identification is limited to simulation, there will be a tendency to choose a smaller sampling period in order to increase the information on the high frequency system. In practice we will choose:

$$\frac{\tau_{ol}}{30} < T < \frac{\tau_{ol}}{10} \tag{4.2}$$

3. Another practical choice for the sampling period is based on the rise time T_{rise} of the step response. Typically, 10 to 15 samples are taken per rise time.

4. Choosing a higher sampling rate than what is actually “appropriate” can drastically increase the “amount” of noise in the measurement.

5. Faster sampling rates can push the discrete-time system to the boundaries of stability.

4.2.2 Outliers detection

An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism. Several among the existing and classical techniques for detecting data outliers can be divided into different approaches, namely, those that make use of statistical models, those based on spatial proximity models and specialized approaches.

4.2.3 Missing data

Sensors can intermittently (missing at random) or systematically undergo failures leading to blank or missing data. This problem can be treated by deletion methods.

4.2.4 Elimination of high frequency disturbances

If the sampled inputs and outputs are corrupted by noise whose spectrum is beyond that of the useful signals, they should be filtered with a low-pass digital filter, for example that of the first order $L\left(q^{-1}\right) = \frac{1-\alpha}{1-\alpha q^{-1}}$:

$$u(k) = \alpha u(k-1) + (1-\alpha)u_e(k) \tag{4.3}$$

$$y(k) = \alpha y(k-1) + (1-\alpha)y_e(k) \tag{4.4}$$

with

$$\alpha = e^{-T/\tau_f} \tag{4.5}$$

Where τ_f represents the filter time constant ($\omega_f = 1/\tau_f$ its cutoff frequency).

4.3 Choice of input signal

In the most general case there is two common input signal that can effectively excite the modes of the system, namely, multi-sinus input and pseudo random binary signal (PRBS). In this section we present the design procedure these two input signals which are considered to be vary practical choice.

4.3.1 Pseudo-random Binary Signal excitation

Pseudo-random signals, generated to approximate white noise, are periodic and perfectly determined (deterministic) signals but which, to an uninformed observer, could appear to evolve under the influence of chance.

As an example, consider a binary pseudo-random signal (PRBS) with period $\theta = 15T$, where T is the sampling period. Figure (4.1) illustrates the auto-correlation function of this pseudo-random signal, which also has a periodicity of $\theta = 15T$. In this example, $\theta = (2^n - 1)\,T$ where $n = 4$ represents the length of the shift register.

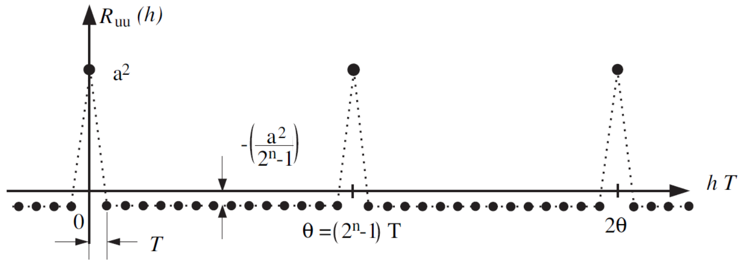

FIGURE 4.1: Auto-correlation function of PRBS in Figure 2.11

When using pseudo-random excitation to identify a dynamic system, it is necessary to verify certain experimental conditions in order to obtain a reliable estimate

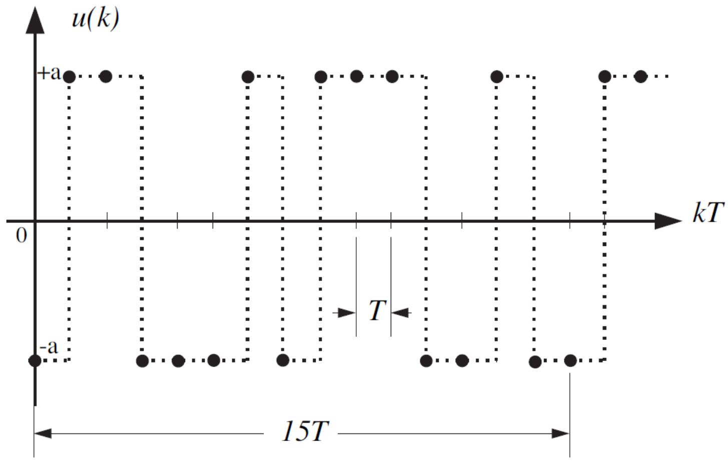

FIGURE 4.2: Pseudo-random binary signal with period $\theta = 15T$.

- It is necessary to provide a period θ greater than the extinction time of the impulse response transients.
- We must choose T sufficiently small and n sufficiently large so that $R_{uu}(h)$ can be assimilated to a impulse of short duration.
- The calculation must be carried out over an integer number of periods θ, that is to say $t_2 - t_1 = p\theta, p$ integer, where t_1 and t_2 represent the start and end of the experiment, respectively.

4.3.2 Generation of pseudo-random binary signals

In this section we provide a simple method for generating an PRBS. These are deterministic binary signals that have white noise like properties (and hence the name pseudo-random) PRBS:

$$u(k) = \text{rem}\left(a_1 u(k-1) + \cdots + a_n u(k-n), 2\right) \tag{4.6}$$

With n coefficients, we can generate $M = 2^n - 1$ full length sequence (zero excluded). The relevant coefficients can be found in the following table for each n

Order	$M = 2^n - 1$	Non-zero indices of $\{a_n\}$
2	3	1,2
3	7	2,3
4	15	1,4
5	31	2,5
6	63	1,6
7	127	3,7
8	255	1, 2, 7, 8
9	511	4,9
10	1023	7,10
11	2047	9,11

TABLE 4.1: Generation of PRBS primates

4.3.3 Multi-sine Input

For a discrete model of the form (3.13) with p parameters to identify, it is possible to choose $u(k)$ as the sum of q sinusoids of distinct frequencies:

$$u(k) = \sum_{i=1}^{q} \sin(\omega_i kT) \tag{4.7}$$

with

$$\begin{cases} q \geq \frac{p}{2} & \text{if } p \text{ is even} \\ q \geq \frac{p+1}{2} & \text{if } p \text{ is odd} \end{cases}$$

To properly identify the parameters, it is therefore necessary to apply a frequency-rich input. In practice, binary excitations (two signal levels) with a variable duration per level are often used. This duration can vary linearly or be generated pseudo-randomly.

4.3.4 Amplitude

The amplitude of the excitation signal represents a compromise between the signal-to-noise ratio and the non-linearity of the system to be identified.

4.3.5 Experiment Duration

Let us consider the case of excitation using a pseudo-random binary signal. Let us denote by T_a the shortest impulse (the clock tick) of the excitation signal with T_a is a multiple of the sampling period:

$$T_a = pT \qquad p = 1, 2, 3, \ldots \tag{4.8}$$

The excitation signal must be conditioned correctly. For example, in order to better identify the static gain of the system with an PRBS, we propose that the duration of the longest impulse is greater than the process establishment time (4.3). We thus obtain:

$$nT_a > 4\tau \tag{4.9}$$

With n the length of the shift register, $n > 5$ so that $R_{uu}(h)$ can be thought of as a unit impulse. On the other hand, to cover the entire spectrum of available frequencies, a test length L is required at least equal to a complete excitation period:

$$L > (2^n - 1)\, T_a \tag{4.10}$$

The relations (4.9) and (4.10) set a lower bound for n and L. For example for $\tau/T = 10$, we obtain:

$$n > \max(5,\ 40/p)$$
$$L > (2^n - 1)\, pT$$

In order to avoid prohibitive test lengths while guaranteeing good identification of low frequencies, we often choose $p = 2.3$ or even 8. For the aforementioned example, we obtain:

$$\text{for } p = 1 \quad n = 40 \quad L > (2^{40} - 1)\, T$$
$$\text{for } p = 8 \quad n = 5 \quad L > (2^5 - 1)\, 8T$$

which represents a reduction in test duration of the order of 2^{32}. However, it is important to note

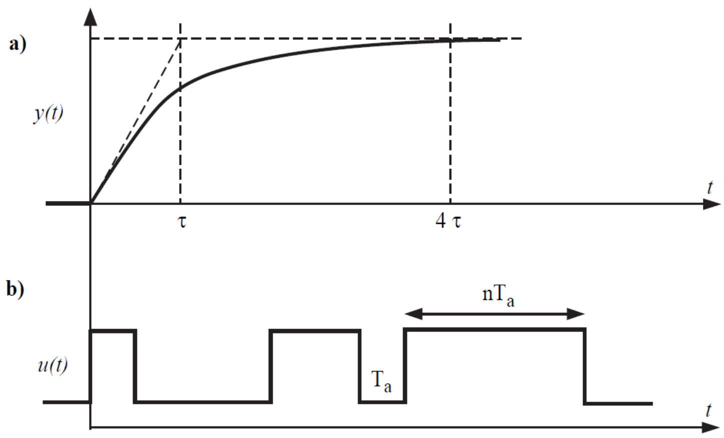

FIGURE 4.3: (a) Index response of a system with its settling time (around 4τ); (b) Longest impulse (nT_a) of an PRBS.

that for $p > 1$ the PRBS no longer has the frequency characteristic of white noise. Indeed, by increasing the richness of PRBS at low frequencies we identify better at low frequencies but we lose precision at high frequencies. This is demonstrated by the spectra of an PRBS with different values of p in figure (4.4).

4.4 Model Validation

Although there are several approaches for model validation (e.g. comparing outputs or Bode diagrams). In this section we rely on statistical methods.

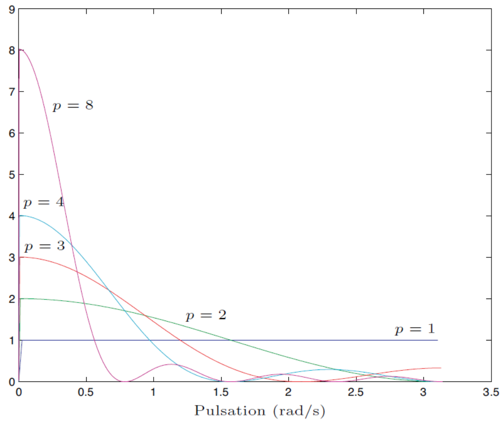

FIGURE 4.4: The spectra of an PRBS with different values of p.

4.4.1 Validation against the expected goal

This method consists of trying and... seeing if the desired goal can be achieved with the identified model. If this is the case, we can consider the model as "good", independently of more formal aspects linked to its development. However, it may prove impossible, too costly or dangerous to validate a model in this way.

4.4.2 Validation of the model with experimental data

Comparing model output with new data

This method consists of using, for the validation step, data not used when establishing the model. This involves comparing the output of the model and the output measured in terms of a quadratic criterion.

Comparing model output with data already in use

This method constitutes a solution to the previous problem, the validation being done with measurements already used to identify the model parameters. However, such an approach is dangerous.

Model comparison in Bode diagram

This method is very useful, because the elements to be compared do not necessarily have the same initial hypotheses: the identified model is of the parametric type, therefore with a prior choice of structure, order, etc., the statement in the diagram de Bode constitutes a non-parametric approach without any a priori structural choice.

4.4.3 Validation by statistical methods

Prediction error whiteness test

If the process and noise models are correct, the prediction error will asymptotically tend towards white noise whose autocorrelation is a Dirac impulse. That is: $R_{\varepsilon\varepsilon}(h) = 0$ for $h \neq 0$ and $R_{\varepsilon\varepsilon}(0) \neq 0$. In practice, $R_{\varepsilon\varepsilon}(h)$ is estimated with the following relation:

$$R_{\varepsilon\varepsilon}(h) = \frac{1}{N-h} \sum_{k=0}^{N-h-1} \varepsilon(k)\varepsilon(k+h) \qquad h = 0, 1, \ldots, N-1 \tag{4.11}$$

Therefore, $R_{\varepsilon\varepsilon}(h)$ will be non-zero for $h \geq 1$ because a finite number of measurements are available and, on the other hand, $\varepsilon(k)$ contains errors structural residuals (order errors, nonlinear effects, etc.). We then consider as a practical validation criterion (with a confidence threshold of 95%)

$$\left| \frac{R_{u\varepsilon}(h)}{\sqrt{R_{\varepsilon\varepsilon}(0)R_{uu}(0)}} \right| \leq \frac{1.96}{\sqrt{N}} \quad \text{for} \quad 0 \leq h < 20, \quad N > 100 \tag{4.12}$$

A correlation between $u(k)$ and $\varepsilon(k+h)$ for negative values of h indicates the presence of feedback from the output to the input, and not a model error.

Testing independence between model output and prediction error

Continuing from the previous point, if the model of the process is correct, the prediction error will be independent not only of the past inputs but also of the outputs of the model y_m. If $y_m(k)$ and $\varepsilon(k)$ are independent, their inter-correlation will be zero, but the unbiased estimate of the inter-correlation function gives:

$$R_{y_m\varepsilon}(h) = \frac{1}{N-h} \sum_{k=0}^{N-h-1} y_m(k)\varepsilon(k+h) \quad h = 0.1, \ldots, \quad N-1 \tag{4.13}$$

with as a practical validation criterion (with a confidence threshold of 95%):

$$\left| \frac{R_{y_m\varepsilon}(h)}{\sqrt{R_{\varepsilon\varepsilon}(0)R_{y_my_m}(0)}} \right| \leq \frac{1.96}{\sqrt{N}} \quad \text{for} \quad 0 \leq h < 20, \quad N > 100 \tag{4.14}$$

This test is important within the framework of the instrumental variables method because it indicates whether the prediction error is indeed uncorrelated with a typical instrumental variable, in this case $y_m(k)$. This test is generic and remains valid even if the instrumental variables were generated differently.

Confidence interval of estimated parameters

The identification of the parameters generally provides an indication of the precision of the results (for example, from the covariance of the parameters (3.103)). It is then possible to generate, with a certain degree of probability, confidence intervals for the identified parameters:

$$\hat{\theta} \in \left[\hat{\theta}_{\min}, \hat{\theta}_{\max}\right] \text{ with 95\%(or 99\%) probability} \tag{4.15}$$

If for a given parameter, this interval includes the value 0, this parameter could very well be neglected and, thus, the structure of the model chosen more simply.

4.5 Parametric identification procedure

The success or otherwise of an identification will depend on the appropriateness of certain choices such as the sampling period, the excitation signal, the delay, the degrees of the different polynomials to model the process and the noise. These choices are made on the basis of a priori knowledge or simple tests carried out on the system to be identified, but they can of course be modified later iteratively if necessary. The parametric identification procedure is summarized as follows

1. Roughly estimate the delay, the dominant time constant of the system and the noise level of the measured signals. This information can be easily obtained, for example using the index response and/or a correlation analysis. It will be possible to deduce an appropriate sampling period and excitation signal (frequency content, amplitude, duration).
2. Estimate the order and delay of the model.
3. Use the ARMAX, OE and BJ structures with these degrees for the process model polynomials.
4. Validate the best models obtained by comparing their output with new data as well as using tests of whiteness of the prediction error and independence between past inputs and the prediction error.

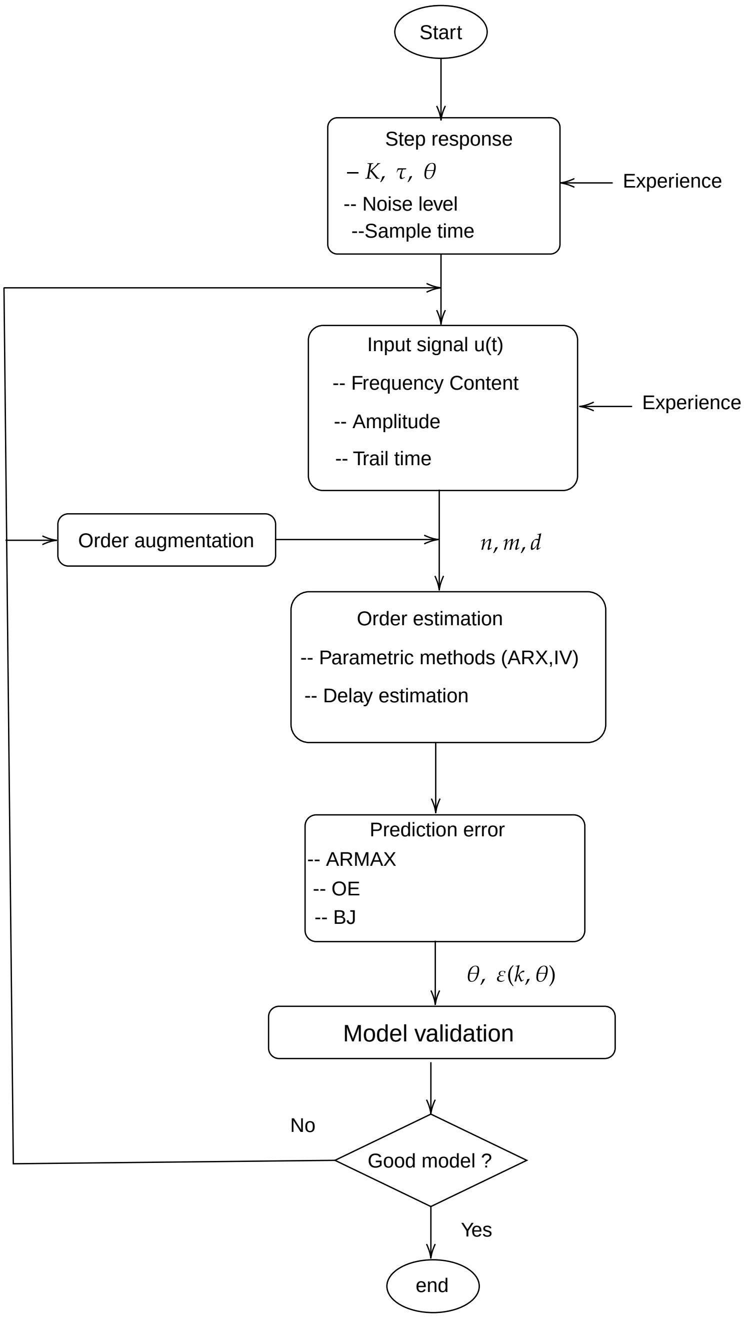

FIGURE 4.5: Parametric identification procedure.

Bibliography

[1] D. Bonvin, D. Bonvin, and D. Bonvin. *Identification de systèmes dynamiques*. Polycopiés de l'EPFL. EPFL, 2000.

[2] Arun K Tangirala. *Principles of system identification: theory and practice*. Crc Press, 2018.