

Model Identification And Data Analysis I

Theory

Christian Rossi

Academic Year 2023-2024

Abstract

The course topics are:

- Basic concepts of stochastic processes.
- ARMA and ARMAX classes of parametric models for time series and for Input/Output systems.
- Parameter identification of ARMA and ARMAX models.
- Analysis of identification methods.
- Model validation and pre-processing.

Contents

1	Introduction	1
1.1	Modeling	1
1.1.1	Error	2
1.1.2	Classification	2
1.2	Estimation problem	3

Introduction

1.1 Modeling

Definition (*System*). A system denoted by S refers to a physical entity designed to convert inputs (causes) into outputs (effects).

Definition (*Model*). A model, symbolized as M , constitutes a mathematical description of a system.

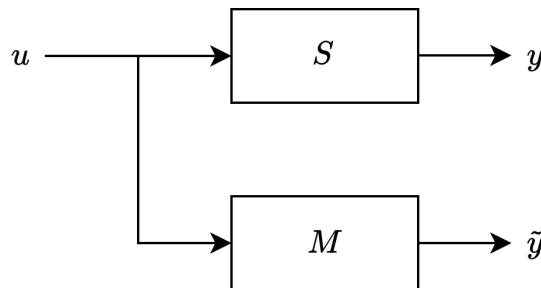


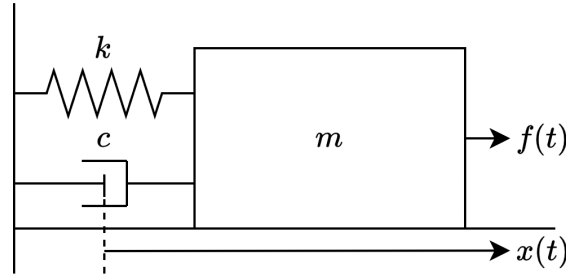
Figure 1.1: Visual representation of system and model

A model can be constructed through various methodologies:

1. *White-box modeling*: this approach relies on established physical laws or existing knowledge. The resultant model is typically generalizable, with clear physical interpretations for each variable. However, precise knowledge of all parameters beforehand is necessary, making it a costly and time-intensive process. Consequently, it's often impractical for complex systems.
2. *Black-box modeling*: this method is based on experimental data. Parameters of the model are estimated using statistical relationships derived from the data. It's feasible even without in-depth knowledge of the underlying processes, and it's comparatively faster and less expensive. However, models generated through this method lack physical interpretability and may not be universally applicable; changes in the system often necessitate repeating the experiment.

Example:

Consider a block with mass m and a spring with an elastic constant k .



In white-box modeling, precise values of parameters such as m , k , and c are required. With this information, the following model can be utilized:

$$m\ddot{x}(t) = f(t) - c\dot{x}(t) - kx(t)$$

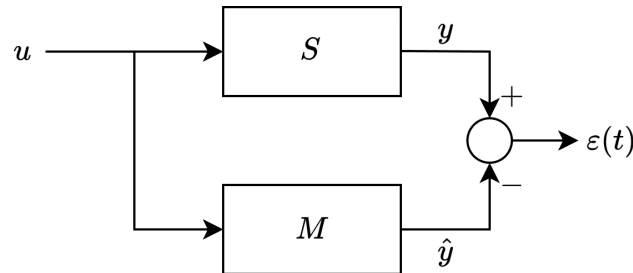
On the other hand, black-box modeling necessitates understanding the relationship between inputs and outputs to derive the model. In this scenario, the model obtained is:

$$x(t) = -a_1x(t-1) - a_2x(t-2) + b_0f(t) + b_1f(t-1) + b_2f(t-2)$$

Here, the parameters are determined from the output-input relationships.

1.1.1 Error

The modeling error, also known as the residual, is calculated as the disparity between the system output and the model output generated with the same input.



When the outputs exhibit similarity based on certain metrics, it signifies that the model accurately mirrors the dynamics of the system. However, if patterns persist within the error graph, it indicates that not all information has been effectively extracted from the data. Conversely, if the error graph lacks of patterns, it is termed as white noise, suggesting an inability to extract further meaningful information from the data. Consequently, a model is deemed complete only when the error demonstrates a completely unpredictable pattern.

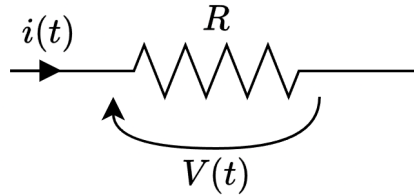
1.1.2 Classification

Static and dynamic A system can be categorized as follows:

- *Static system*: in this type of system, knowledge of the input variables alone is adequate to determine the output value. Classical machine learning primarily addresses the black-box modeling of static systems.
- *Dynamic system*: this refers to a system with memory, wherein the past behavior of the output impacts its current value.

Example:

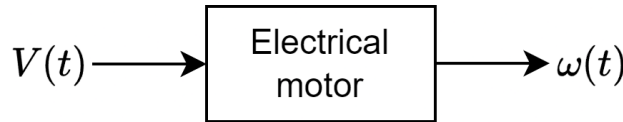
An illustration of a static system is represented by a circuit containing a resistor.



This system solely relies on the voltage across the resistor at each moment, adhering to Ohm's law:

$$i(t) = \frac{V(t)}{R}$$

On the other hand, an example of a dynamic system is exemplified by an electrical motor, wherein under certain conditions, even if the input remains constant, the output persists in its evolution.



Discrete and continuous Systems can be further categorized based on their time description, which can be either discrete or continuous. Natural and physical phenomena are inherently continuous and are often mathematically described using ordinary differential equations. On the other hand, discrete systems are mathematically described using difference equations.

However, a computer can only handle a limited amount of data. However, computers have limitations in handling data, which necessitates the sampling of signals at discrete intervals with a sampling time T_s . This ensures that only a finite amount of data is stored at discrete time points $t \cdot T_s$, where $t = 1, \dots, N$:

$$y(t) = y(t \cdot T_s)$$

1.2 Estimation problem

An estimation problem involves quantifying an unknown parameter through estimation. This parameter, denoted as ϑ , can be discrete or continuous, scalar or vectorial, and constant or time-variant. We are provided with a set of observations, d , taken at various time points t_1, t_2, \dots, t_N , formally defined as:

$$d = \{d(t), t \in T\}$$

Our goal is to derive an estimator to obtain an estimate of the unknown variable ϑ , expressed as:

$$\vartheta = f(d)$$

Definition (*Estimate*). An estimate, denoted as $\hat{\vartheta}$, is a value produced by an estimator and depends on the input values of the estimator.

For a parameter ϑ with a constant value, we seek the estimate $\hat{\vartheta}$. For a parameter $\vartheta(t)$ with a dynamic value, we aim to find the estimate $\hat{\vartheta}(t|t_N)$ where the value of t_N is provided. The choice of t determines the nature of the estimation:

- *Prediction*: when $t > t_N$, indicating a time instant beyond t_N , we are forecasting a future event.
- *Filtering*: when $t = t_N$, we are estimating the noise in the estimator.
- *Regularization* or *interpolation* or *smoothing*: when $t < t_N$, representing a time instant before t_N , we are estimating variables that are not directly accessible.