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

CHAPTER 1

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.

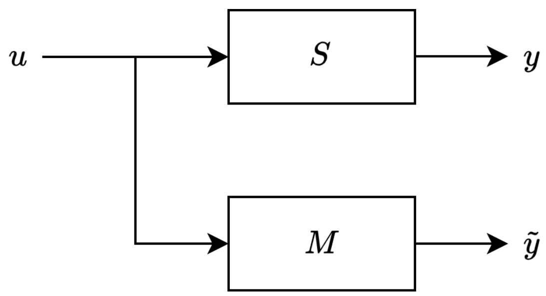


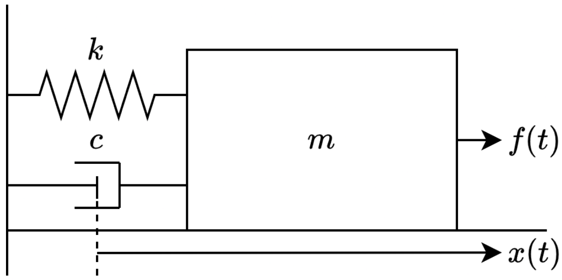
Figure 1.1: Visual representation of system and model

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:

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

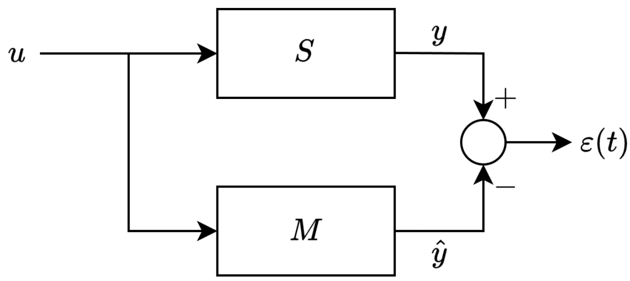
On the other hand, black-box modeling necessitates understanding the relationship between inputs and outputs to derive the model. In this scenario, the model obtained is:

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

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.

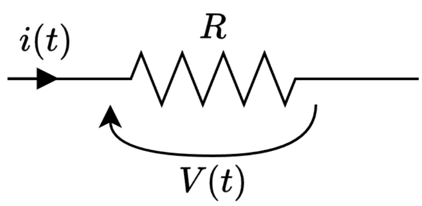
1.1.2 Classification

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.

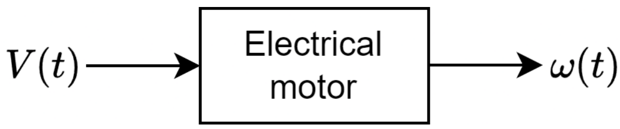
Example:
An illustration of a static system is represented by a circuit containing a resistor.



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 $\widehat{\vartheta}$, 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 $\widehat{\vartheta}$. For a parameter $\vartheta(t)$ with a dynamic value, we aim to find the estimate $\widehat{\vartheta}(t|t_N)$ 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.

CHAPTER 2

Stochastic processes

2.1 Prediction problem

To predict the value of $v(t)$ given a set of observations $\{v(1), v(2), \dots, v(t-1)\}$, we can devise a predictor using the formula:

$$\widehat{v}(t|t-1) = f\left(v(t-1), v(t-2), \dots, v(1)\right)$$

In this formulation, we impose certain constraints:

- The function f is linear.
- Older data have diminishing importance compared to recent ones (finite memory predictor).
- The function remains invariant over time.

Under these simplifications, the predictor takes the form:

$$\widehat{v}(t|t-1) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n)$$

Here, v is represented as a vector:

$$v = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}^T$$

A reliable prediction is one that yields estimates closely aligned with the actual values. The accuracy of these estimates hinges on the parameters a_i in the v vector. Determining these parameter values equates to identifying the model that best characterizes the data distribution. This task translates into an optimization problem.

2.1.1 Model quality

Uncertainty represents a critical aspect of noise in prediction problems, yet its precise magnitude cannot be predetermined. The sole method to calculate an estimate of uncertainty is by comparing known values with those provided by the estimator at corresponding instants.

Example:
Consider the predictor given by:

$$\widehat{v}(t|t-1) = a_1 v(t-1) + a_2 v(t-2) + a_3 v(t-3)$$

The values to be examined are:

- $\widehat{v}(4|3) = v(1)v(2)v(3)$: to be compared with $v(4)$.
- $\widehat{v}(5|4) = v(2)v(3)v(4)$: to be compared with $v(5)$.

After comparing all possible sequences, we can generate a sequence of residuals using the formula:

$$\varepsilon(i) = v(i) - \widehat{v}(i|i-1) \qquad i = n+1, \dots, N$$

From this sequence, we seek to find v by minimizing the following function:

$$\mathcal{J}(v) = \sum_{n+1}^{N} \varepsilon(i)^2$$

It's worth noting that the error is squared to ensure it is always considered as positive. A predictor is considered effective if the remaining error exhibits no discernible pattern, indicating that any remaining error is attributable solely to white noise.

Definition (*White noise*)**.** White noise refers to an error characterized by its lack of correlation between values at different points in time.

Consequently, if the residual is white noise, it signifies that there's no meaningful information within it that can enhance predictions.

Finally, to derive the accurate value from an estimate, we must incorporate the residual into the estimate. Thus, the previous formula transforms to:

$$\widehat{v}(t) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n) + \varepsilon(t)$$

This implies that addressing the prediction problem involves examining a stochastic system.

2.1.2 Zeta transform

The same system can be reformulated using the Z-transform defined as:

$$V(z) = \mathcal{Z}\left[v(t)\right]$$

When considering the same system at a time $t-1$ in the Z-transform formulation, we have:

$$\underbrace{z^{-1}}_{\text{unity delay operator}} \cdot V(z) = \left[v(t-1)\right]$$

By incorporating the system described above with the time-domain equation:

$$\widehat{v}(t) = a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n) + \varepsilon(t)$$

It's feasible to rewrite the same model in the frequency domain with the Z-transform, resulting in:

$$V(z) = \mathcal{Z}\left[a_1 v(t-1) + a_2 v(t-2) + \cdots + a_n v(t-n) + \varepsilon(t)\right]$$

This simplifies to:

$$V(z) = a_1 z^{-1} V(z) + a_2 z^{-2} V(z) + \cdots + a_n z^{-n} V(z) + \xi(z)$$

Note that this formula can also be expressed using operatorial notation:

$$V(z) = a_1 z^{-1} v(t-1) + a_2 z^{-2} v(t-2) + \cdots + a_n z^{-n} v(t-n) + \xi(t)$$

Rearranging terms such that all elements multiplied by $V(z)$ are on the left side and the residual $\xi(z)$ is on the right side, we obtain:

$$V(z)\left(1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_n z^{-n}\right) = \xi(z)$$

Finally, from this expression, we can derive the transfer function as the ratio of $V(z)$ to the residual $\xi(z)$:

$$\frac{V(z)}{\xi(z)} = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_n z^{-n}}$$

2.1.3 Summary

The components of an identification problem consist of:

- A system $\mathcal{S}$ requiring modeling.
- A model $\mathcal{M}$ to be ascertained, describing the system.
- An identification algorithm $\mathcal{J}$ governing data processing.
- An identification experiment $\mathcal{E}$ providing the data.

From these elements, it's crucial to emphasize that the model cannot convey more information than what is inherent in the data.